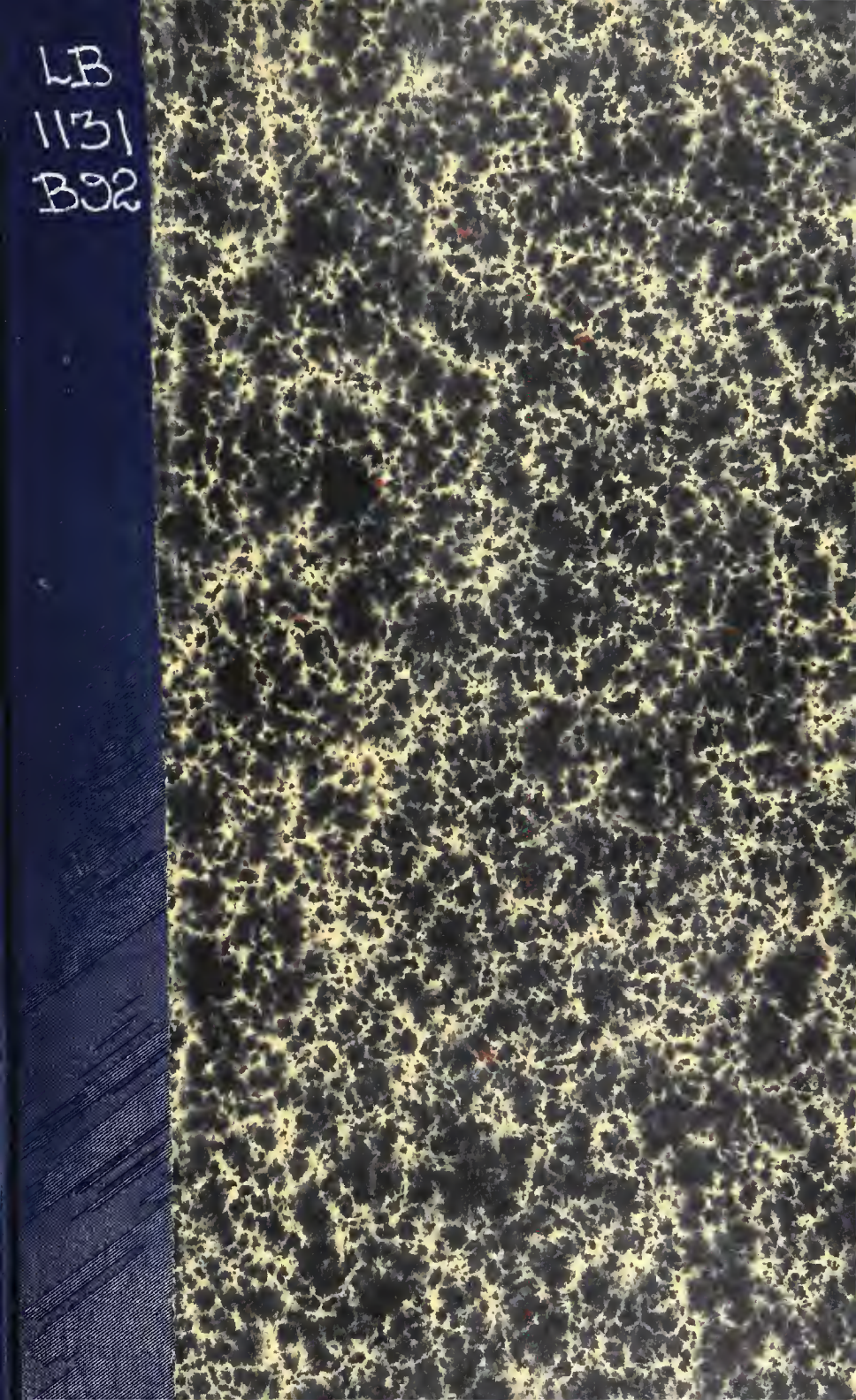


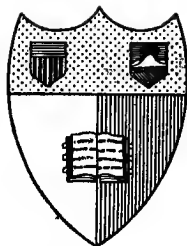
LB
1131
B92



LB

1131

B92



Cornell University Library

Ithaca, New York

THE GIFT OF

*Columbia University
(in exchange)*

The date shows when this volume was taken.

To renew this book copy the call No. and give to
the librarian.

HOME USE RULES

All Books subject to recall

All borrowers must register in the library to borrow books for home use.

All books must be returned at end of college year for inspection and repairs.

Limited books must be returned within the four week limit and not renewed.

Students must return all books before leaving town. Officers should arrange for the return of books wanted during their absence from town.

Volumes of periodicals and of pamphlets are held in the library as much as possible. For special purposes they are given out for a limited time.

Borrowers should not use their library privileges for the benefit of other persons.

Books of special value and gift books, when the giver wishes it, are not allowed to circulate.

Readers are asked to report all cases of books marked or mutilated

Do not deface books by marks and writing.

NOV 9 1953 HV

JAN 8 1954 IS

MAR 21 53

APR 25 1961 JR

Cornell University Library

LB1131 .B92

Educational diagnosis of individual pupils



3 1924 032 705 497

olin

This book was digitized by Microsoft Corporation in cooperation with Cornell University Libraries, 2007.

You may use and print this copy in limited quantity for your personal purposes, but may not distribute or provide access to it (or modified or partial versions of it) for revenue-generating or other commercial purposes.

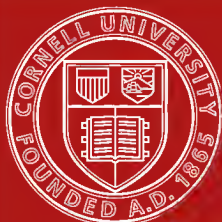
Educational Diagnosis of Individual Pupils

A Study of the Individual Achievements of Seventy-Two Junior High School Boys in a Group of Eleven Standardized Tests

By
CHESTER A. BUCKNER

Submitted in partial fulfillment of the requirements for the Degree
of Doctor of Philosophy in the Faculty of Philosophy,
Columbia University

Published by
Teachers College, Columbia University
NEW YORK CITY
1919



Cornell University
Library

The original of this book is in
the Cornell University Library.

There are no known copyright restrictions in
the United States on the use of the text.

<http://www.archive.org/details/cu31924032705497>

Educational Diagnosis of Individual Pupils

A Study of the Individual Achievements of Seventy-Two Junior High School Boys in a Group of Eleven Standardized Tests

By
CHESTER A. BUCKNER

Submitted in partial fulfillment of the requirements for the Degree
of Doctor of Philosophy in the Faculty of Philosophy,
Columbia University

Published by
Teachers College, Columbia University
NEW YORK CITY

1919

56
CORNELL
UNIVERSITY
LIBRARY

LB
1131
B92

A490715

Copyright, 1919, by CHESTER A. BUCKNER

ACKNOWLEDGMENTS

FOR the use of data secured and for the privilege of securing additional data concerning the achievements of certain pupils in the Speyer School of Teachers College, I am indebted to Professor Thomas H. Briggs. I am grateful to the teachers of this school and to Dr. E. K. Fretwell for their coöperative aid in administering the tests.

For the suggestion of the field of research and for helpful supervision of the work my obligation to Professor Briggs is also gratefully acknowledged. To Professor George D. Strayer I am likewise indebted for constructive criticism.

Because of her devoted interest and untiring assistance in the prosecution of this study I am most grateful to my wife, Neva Starrett Buckner.

C. A. B.

CONTENTS

SECTION	PAGE
I. THE PROBLEM	1
II. PRELIMINARY INVESTIGATION	3
III. EXPERIMENTAL MATERIAL AND METHOD.....	10
1. The Subjects	10
2. The Administration and Scoring of the Tests.....	12
3. Special Testing	17
IV. STATISTICAL TREATMENT	19
1. Transmutation and Distribution of Scores.....	19
2. The Use of Averages and Variabilities.....	21
3. Redistribution of Scores	28
V. INDIVIDUAL VARIABILITY COMPARED WITH GROUP VARIABILITY	30
1. The Amount of Individual Variability.....	30
2. Distribution of Individual Variability.....	37
3. Overlapping of Divisions of the Group.....	46
VI. EXTREME VARIABILITY IN INDIVIDUAL CASES.....	51
1. Extreme Variability in Different Tests.....	51
2. Extreme Variability of Different Boys.....	55
3. Reduction of Variability by Re-examination.....	60
4. The Causes of Extreme Variability.....	63
VII. CORRELATION BETWEEN MEASURES OF ABILITY AND VARIABILITY	75
1. Correlation Between Measures of Ability.....	75
2. Correlation Between Measures of Variability.....	78
3. Correlation Between Measures of Ability and Variability	79
VIII. CONCLUSIONS	81
APPENDIX	85

INDEX OF TABLES

NUMBER	PAGE
I. The 34 Most Erratic Scores Distributed by Quartiles according to the Different Classifications.....	5
II. The 34 Most Erratic Scores of Each Classification Compared according to the Number Above and the Number Below the Median Score of the Individual.....	6
III. Distribution of the 34 Most Erratic Scores among the Tests according to the Different Classifications.....	6
IV. Summary of Variation in Ranks of the 97 Pupils in the Eleven Tests	7
V. Distribution of Ranges in Ranks of the 97 Pupils in the Eleven Tests	7
VI. Distribution of Scores of the 97 Pupils in the Eleven Tests according to S.D. Distance From the Median Score of the Individual	8
VII. Distribution of the 72 Boys among Groups in School....	11
VIII. The Tests and the Times at Which They Were Given.....	18
IX. The Semi-Interquartile-Range (Q) of the Distribution of the Original Scores for Each Test.....	20
X. Distribution of Scores Transmuted into Multiples of Q Above and Below the Median of the Original Distribution in Each Test	20
XI. Average of the Individual Semi-Interquartile-Ranges in the Eleven Tests	31
XII. Distribution of the Individual Semi-Interquartile-Ranges (Approximation) in the Eleven Tests.....	32
XIII. Averages in Connection with Individual Ranges in Scores Transmuted into Multiples of Q.....	33
XIV. Comparison of the Variability of Individual Ranges in the Eleven Tests by Quartiles and by Tertiles.....	35
XV. Distribution (in Per Cents) of Scores Above and Below the Individual Medians in the Eleven Tests Transmuted into Multiples of Q by the Original Distributions	38
XVI. Distribution (in Per Cents) of Scores Above and Below the Individual Medians in Certain Tests Transmuted into Multiples of Q by the Original Distributions....	40
XVII. The Q of the Distribution of Scores Above and Below the Individual Medians for the Three Testings and for Certain Tests Combined	41
XVIII. Distribution by Tertiles of Ranges Above and Below the Individual Medians in the Eleven Tests in Values of Q	42
XIX. Distribution by Tertiles of Ranges Above and Below the Individual Medians in Certain Tests in Values of Q	43
XX. Per Cent of Scores in Each Quartile Above the 75 Percentile and the Median of Each Quartile Higher, and the Per Cent in Each Quartile Below the Median and the 25 Percentile of Each Quartile Lower.....	47
XXI. Difference in Achievement Between Quartiles Measured in Terms of the Q Variability of the Group.....	48

Index of Tables

NUMBER		PAGE
XXII.	Number of Scores in the Different Tests 3 Q or More Plus or Minus	52
XXIII.	Number of Scores 3 Q or More Plus or Minus by Tertiles and Total for Each Testing.....	55
XXIV.	Number of Boys Having Scores 3 Q or More Plus or Minus in Each Type of Test in Either One or More Testings	56
XXV.	Number of Boys Making Different Numbers of Scores 3 Q or More Plus or Minus in All Three of the Testings	56
XXVI.	Number of Boys Making Scores 3 Q or More Plus or Minus and the Number of Scores of Either Type that Each Boy Made	58
XXVII.	Number of Boys Making Scores 3 Q or More Plus, Minus, and Plus and Minus in One or More of the Three Testings	59
XXVIII.	Number of Boys Having Scores 3 Q or More Either Plus, or Minus, or Plus and Minus by Tertiles and Total for Each Testing.....	60
XXIX.	Comparison of Scores in Original and Special Tests of Certain Boys Having Scores 3 Q or More Minus in Original Tests.....	61
XXX.	Comparison of Scores of Certain Boys in Special Tests with Their Scores in Corresponding Original Tests....	63
XXXI.	The Amount of Q Representing One-Half the Interval of the Distributions of the Different Tests.....	64
XXXII.	The Variability of Certain Individuals in the Ranking of Their Own Achievements for the Different Testings	70
XXXIII.	Teachers' Ratings on Certain Points Concerning the Work of Fifteen Pupils	71
XXXIV.	Correlation Between Composite Rankings in Ability.....	76
XXXV.	Correlation Between Measures of Variability in the Eleven Tests at the Different Times They Were Given.....	78
XXXVI.	Correlation Between Measures of Ability and Variability in the Eleven Tests.....	79
XXXVII.	Distribution of Scores Above and Below the Individual Medians in the Eleven Tests Transmuted into Multiples of Q by the Original Distributions.....	87
XXXVIII.	Distribution of Scores Above and Below the Individual Medians in Certain Tests Transmuted into Multiples of Q by the Original Distributions.....	88
XXXIX.	Original Scores by Tests and by Individuals. February, 1916	89-90
XL.	Original Scores by Tests and by Individuals. February, 1917.....	91-92
XLI.	Original Scores by Tests and by Individuals. June, 1917	93-94
XLII.	Original Scores by Tests and by Individuals. Additional Tests	95

INDEX OF FIGURES

NUMBER	PAGE
1. Different Forms of Distribution of the Scores of Individuals in the Eleven Tests.....	4
2. Distribution of the Scores, in Values of Q, of Three Individuals. February, 1916 Tests.....	24
3a, b, c to 13a, b, c. Distribution of Scores in Each Test Transmuted into Multiples of Q Above and Below the Median of the Original Distribution	26-27
14. Showing the Effect of Two Different Forms of Distribution upon the Q	28
15. Chart Showing the Value in Q of Each Score of the June, 1917 Tests. (Insert opposite page 28).....	28
16a to 17c. Distribution (in Per Cents) of Scores Above and Below the Individual Medians in the Eleven Tests Transmuted into Multiples of Q by the Original Distributions.....	39
18 to 21. Distribution (in Per Cents) of Scores Above and Below the Individual Medians in Certain Tests Transmuted into Multiples of Q by the Original Distributions.....	41
22a to 23c. Distribution by Tertiles of Ranges Above and Below the Individual Medians in the Eleven Tests and in the Eight Trabue Tests in Values of Q.....	44
24a to 25c. Distribution by Tertiles of Ranges Above and Below the Individual Medians in the Six Mathematics and Five Directions Tests	45

EDUCATIONAL DIAGNOSIS OF INDIVIDUAL PUPILS

I

THE PROBLEM

EDUCATIONAL diagnosis presents many problems, each with its specific implications. In an approach to the present study it does not seem imperative to consider either a logical organization of these problems or an exhaustive summary of relevant investigations; however, mention of a few problems and methods will assist in the orientation of this study in the general field. One phase of educational diagnosis is based upon standardized tests and scales, which have been used to study and compare the attainments of groups of pupils, usually school grades or school systems. The average or median achievement of different groups, the extent of overlapping, the amount of variability, and the distribution of results have been used as measures for comparison. The relation of the attainment of a group in one function to its attainment in another function or trait has been studied extensively and expressed by various formulæ for correlation. Results obtained by standardized measurements have been compared with teachers' judgments, directly by teachers' rankings and indirectly by comparison with school marks. Tests of the same function or trait have been compared and ranked as to merit. Tests of different traits have been compared and ranked as to their merit in the evaluation of general intelligence. Mistakes made most frequently by the group have been studied. These examples of the use of standardized tests suggest that the trend of the movement in scientific measurements has been to emphasize the group.

In these studies and in even more extensive ones now being undertaken, the individual even though not lost sight of has not received as much attention as the group. It would seem that greater emphasis should be placed upon the measurement of the individual and the interpretation of results along with the measurement of the group and the further development of the instrumentalities of measurement. It is the purpose of this study to ascertain to what extent and with what degree of rela-

bility standardized tests and scales can be used to discriminate educational attainments of the individual. Is it possible to diagnose a case and to prescribe specific mental work on the basis of achievements in such tests? The following are some of the questions to be considered.

1. How can individual measures of achievement in different tests be compared or equated without losing the refinement of the original scores?

2. Do scores of equal value in a given test necessarily have the same meaning for two or more individuals?

3. What is the amount of the individual's variability among the different tests?

4. How are the scores of the individual distributed with respect to some measure of his central tendency?

5. How do the bright, the mediocre, and the dull pupils compare with each other in their variability and distribution of achievements?

6. To what extent are there extremely variable or erratic scores?

7. How do the bright, the mediocre, and the dull pupils compare in the number of erratic scores which they make?

8. What are the causes of the extremely variant scores?

9. What is the relation between different measures of ability, between different measures of variability, and between measures of ability and variability?

The specific purpose of this investigation is to determine the individual achievements of seventy-two junior high school pupils in a group of eleven tests given at three different times during a period of a year and a half. The tests have been used to rank these pupils in achievement, to determine the amount of variability of the group in a single test, and to determine the amount of variability of the individual in the eleven tests. That the data obtained from these tests are valuable in the general direction of the work of these pupils has been demonstrated at the Speyer School of Teachers College. That such data can be used to advantage in the prescription of special work in certain cases is a logical assumption. This, however, should be tested by practice and by further experimentation.

II

PRELIMINARY INVESTIGATION

THE purpose of this section of the study is to answer the first question proposed in the statement of the problem, namely: How can individual measures of achievement in different tests be compared or equated without losing the refinement of the original scores? This section is introduced not only to describe a way of equating measures but also to compare two methods of equating measures of achievement in different tests and to determine by which method more reliable results can be obtained. Special emphasis is placed upon the classification of extremely variable or erratic scores.

The data for the preliminary investigation consist of the scores of ninety-seven seventh grade boys in eleven standardized tests and scales given in February, 1916. The tests are: Woody Arithmetic Scales, Series A, Multiplication and Division; Trabue Completion-Test Language Scales, Scale B and Scale C; Thorndike Reading Scale Alpha 2, Part II; Thorndike Reading Scale A, Visual Vocabulary; Composition, scored by the Hillegas Scale for the Measurement of Quality in English Composition; Ayres Measuring Scale for Ability in Spelling; Woodworth and Wells Association Tests, Opposites, Mixed Relations, and Easy Directions. The description of the subjects, the tests, and the scoring of the tests in Section III, Experimental Material and Method, is applicable here and is omitted from this section because the chief concern here is the evaluation of methods of statistical treatment.

The first method used to compare individual measures of achievement will be called Classification by Rank. By this method the scores of each test were arranged in frequency tables according to the original scores of the papers. The scores were then turned into ranks. The highest score was ranked one and the lowest score was ranked ninety-seven. In cases of tied scores the mid-rank of the interval was given to each score. The eleven ranks of each individual were assembled and arranged in order from highest to lowest. The rank by the original distribution was retained. The variability of a score was then measured

by its distance, in terms of ranks by the original distribution, from the median rank of the individual. Obviously the ranks of an individual could to varying degrees approximate three forms of distribution,—a distribution skewed downward from the median, a distribution skewed upward from the median, and a distribution approximating the normal surface of frequency.

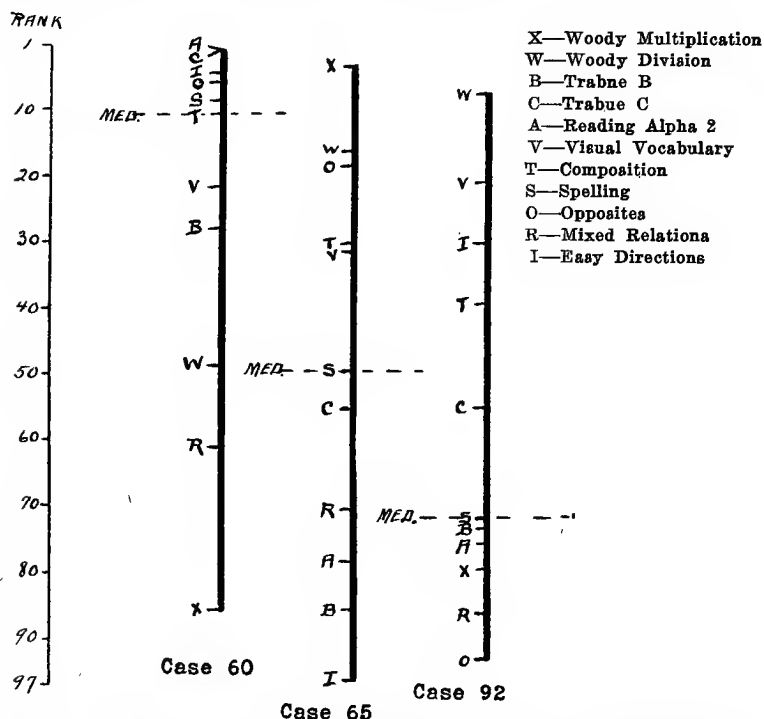


Fig. 1. Different Forms of Distribution of the Scores of Individuals in the Eleven Tests.

The cases in Fig. 1 illustrate these forms. These are actual cases selected from the group of ninety-seven boys. The scale at the left of the plate represents the range in ranks which could be obtained in each test. The letters refer to the tests in which the ranks indicated were made. The case numbers 60, 65, and 92 are the serial numbers which these boys chanced to have when the names of the group were arranged in alphabetical order. These are extreme cases, but only in the sense that they are near the limit of the range of the respective forms which they are selected to illustrate, and not in the sense that they are markedly different from other cases.

Results obtained from this method are shown in Tables IV and V. These results will be discussed in connection with the results from the other method.

The second method will be called Classification by Standard Deviation (S.D.).¹ It is like the first method only to the point of the frequency tables of the original scores. Using these tables the original scores were transmuted into multiples of S.D. The scores of each pupil in multiples of S.D. of the original distribu-

TABLE I
THE 34 MOST ERRATIC SCORES DISTRIBUTED BY QUANTILES ACCORDING
TO THE DIFFERENT CLASSIFICATIONS

	Quartile			
	I	II	III	IV
Classification by Rank	9	6	9	10
Classification by S.D.	4	8	6	16

tions were then collected and arranged in order of S.D. value from highest to lowest. The variability of a score was then measured by its distance, in terms of S.D. by the original distribution, from the median score of the individual. The distribution of all the scores in multiples of S.D. is given in Table VI. The results show that there are thirty-four scores which deviate from the medians of the respective individuals by more than 2 S.D.

These two methods of equating scores and determining individual variability will now be compared in order to arrive at some basis for choosing the one which will produce the more reliable results. In Table I the thirty-four most erratic scores in each classification are distributed among the quartiles of the group, the quartiles being determined from the median ranks of the individuals. Quartile I is the highest. The table reveals a rather marked difference between the two classifications. By the Classification by Rank the erratic scores are distributed quite evenly among the four quartiles. The Classification by S.D. produces decidedly the greatest number of erratic scores in Quartile IV, while it produces relatively few in Quartile I.

Another way of comparing these methods is by dividing the 34 most erratic scores of each classification into the number above the median and the number below the median score of the individual. This is done in Table II. By this comparison

¹ Mean Square Deviation.

Educational Diagnosis of Individual Pupils

Classification by S.D. differs very decidedly from the Classification by Rank. By S.D. the number of erratic scores below the median greatly exceeds the number above the median; by Rank the number of scores above and the number of scores below the median are about equal.

TABLE II

THE 34 MOST ERRATIC SCORES OF EACH CLASSIFICATION COMPARED
ACCORDING TO THE NUMBER ABOVE AND THE NUMBER BELOW
THE MEDIAN SCORE OF THE INDIVIDUAL

	<i>Above the Median</i>	<i>Below the Median</i>
Classification by Rank	18	16
Classification by S.D.	4	30

That these two methods do not affect the same scores in a different way, as might be inferred from Table II, is shown by the fact that of the 34 most erratic scores in each classification only eight are common to both classifications.

The last method of comparing the classifications directly was distributing the 34 most variable scores among the eleven tests. The results are brought together in Table III. Here also there are some rather marked differences between the two classifications. The greatest contrast in the number of erratic scores is found in the case of spelling.

TABLE III

DISTRIBUTION OF THE 34 MOST ERRATIC SCORES AMONG THE TESTS
ACCORDING TO THE DIFFERENT CLASSIFICATIONS

<i>Classification by</i>	<i>Woody Multiplication</i>	<i>Woody Division</i>	<i>True B</i>	<i>True C</i>	<i>Reading Alpha 2</i>	<i>Visual Vocabulary</i>	<i>Composition</i>	<i>Spelling</i>	<i>Opposites</i>	<i>Mixed Relations</i>	<i>Easy Directions</i>
Rank	5	7	5	2	1	1	5	2	2	3	1
S.D.	7	6	3	0	1	0	2	8	3	2	2

Indirectly further comparison of the two classifications can be made by a study of Tables IV, V, and VI. Table IV shows that the average range of pupils who stand highest and of those who stand lowest is considerably less than that of those of average ability. According to range in ranks above and below the median score of the individual the four quartiles have close in-

verse relation. The average range above the median of the first quartile is approximately the same as the average range below the median of the fourth quartile. The inverse relation holds right through to the average range above the median of the fourth quartile which is approximately the same as the average range below the median of the first quartile. The range between the last two scores at each end of the distribution of the individual's scores has an inverse relation similar to that of the average range above and below the median. In average S.D. the quartiles have about the same relation as in average range.

TABLE IV
SUMMARY OF VARIATION IN RANKS OF THE 97 PUPILS IN THE
ELEVEN TESTS

			Average Range		Av. Interval between Last Two Scores		
			Av. Range	Above Med.	Below Med.	Above Med.	
Quartile	I	69.8	23.3	46.6	17.8	3.2	22.2
"	II	81.6	36.2	45.4	13.1	8.2	26.9
"	III	78.5	44.7	33.8	6.6	10.9	26.4
"	IV	67.6	45.9	21.8	2.3	15.6	22.2

Table V is a distribution of the ranges in ranks. From this the median range in ranks, according to the original distribution, is found to be 76.6 which is 82 per cent of the maximum possible range. Taken at its absolute value this seems to be a high per cent. Whether or not its relative value is high must await further investigation. This question will be considered further in connection with Table XIII in Section V.

TABLE V
DISTRIBUTION OF RANGES IN RANKS OF THE 97 PUPILS IN THE
ELEVEN TESTS

Value in Ranks	Frequency	Value in Ranks	Frequency
90 to 94	8	60 to 64	8
85 to 89	17	55 to 59	5
80 to 84	14	50 to 54	2
75 to 79	15	45 to 49	
70 to 74	13	40 to 44	1
65 to 69	12	35 to 39	
		30 to 34	2

Table VI is a distribution of the total number of scores according to their S.D. distance, in terms of the original distribution, from the median score of the individual. Its chief significance lies in the fact that there are more extreme or erratic scores below the median than there are above, and also that they are more widely scattered.

TABLE VI
DISTRIBUTION OF SCORES OF THE 97 PUPILS IN THE ELEVEN TESTS
ACCORDING TO S.D. DISTANCE FROM THE MEDIAN SCORE
OF THE INDIVIDUAL

<i>Value in S.D.</i>	<i>Frequency</i>	<i>Value in S.D.</i>	<i>Frequency</i>
+2.6 to +3.0	3	Med. to — .5	287
+2.1 to +2.5	1	— .6 to —1.0	116
+1.6 to +2.0	21	—1.1 to —1.5	69
+1.1 to +1.5	62	—1.6 to —2.0	32
+ .6 to +1.0	137	—2.1 to —2.5	16
Med. to + .5	309	—2.6 to —3.0	3
		—3.1 to —3.5	2
		—3.6 to —4.0	1
		—4.1 to —4.5	3
		—4.6 to —5.0	2
		—5.1 to —5.5	2
		—5.6 to —6.0	
		—6.1 to —6.5	1

With the results of the preliminary investigation at hand the question was to decide which method of classifying an individual's scores should be used to secure the more reliable results.

The Classification by Rank does not show as much evidence of reliability for a study of this kind as a classification by some measure of variability. A significant weakness,—the one which is the reason for its elimination in an investigation of this type of problem—is the fact that unequal differences between original scores cannot be discriminated. By the frequency tables of the original scores two items, one in each of two contiguous intervals in one table are ranked consecutively; likewise two other items of the same table, although one is removed several intervals from the other, are ranked consecutively if there are no items intervening. In studies where the scores near the central tendency are of chief concern this defect is not of so much importance, but in this study it seems to make the method unreliable.

The Classification by S.D. overcomes the weaknesses pointed out in connection with the other one. The value taken for unity is the amount of variability which is "one of the most constant things about a variable fact" ¹ Different magnitudes of the original frequency tables are preserved because they are expressed in percentages of a constant. Therefore the Classification by S.D. appears to be the superior method.

After the method of Classification by S.D. was chosen as being the more reliable, and the work of this study by a similar method was well under way further confirmation of its reliability was discovered. This will be discussed in Section IV.

The first question asked in Section I concerning the comparing or equating of measures of achievement of the individual in different tests has now been answered. Two measures, Classification by Rank and Classification by S.D., have been described and the results produced by each have been compared. The Classification by S.D. shows marked superiority over the other classification. Its use seems to produce more reliable results,—results by which individual achievements can be compared or equated still retaining practically all the refinement of the original scores.

¹ Trabue, M. R., *Completion-Test Language Scales*, Teachers College, Columbia University, Contributions to Education, No. 77, p. 30.

III

EXPERIMENTAL MATERIAL AND METHOD

1. THE SUBJECTS

The subjects for this investigation were a group of boys in the Speyer School of Teachers College, Columbia University. There were seventy-two individuals for whom complete records were secured in all eleven tests in all three testings, February 1916, February 1917, and June 1917. This school was opened as an experimental academic junior high school in February 1916. The group which entered first, about two hundred in all, came from twenty-four classes in five of the public schools in New York City, Nos. 5, 10 B, 43, 184, and 186. The seventy-two subjects for this study were among this group.

Before the experimental school was opened the twenty-four classes in the public schools were given the following tests: Woody Arithmetic, Multiplication Scale, Series A; Trabue Completion-Test Language Scales B and C; Composition, scored by the Hillegas Scale; and Ayres Spelling Scale. Soon after they entered Speyer School six additional tests were given: Woody Arithmetic, Division Scale, Series A; Thorndike Reading Scale Alpha 2, Part II; Thorndike Reading Scale A, Visual Vocabulary; Woodworth and Wells Association Tests, Opposites, Mixed Relations, and Easy Directions. Complete records of the scores in all of the eleven tests were secured for ninety-seven of the boys entering. These tests will be referred to throughout this study as the February 1916 tests.

For purposes of this investigation it is important to know whether the boys are a highly selected group or whether they represent the different grades of ability in typical classes beginning the seventh school grade. Studying these same boys in connection with a different problem Dr. E. K. Fretwell answers this question as follows: "It is noted then that the Speyer group is, on the basis of achievements in these five tests, somewhat better than the other group, though only slightly better. It should also be pointed out that this group coming to Speyer

did not cluster around the median of achievement and that there were all kinds of pupils from the brightest to very nearly the dullest. On this point the estimates of the twenty-four teachers are in accord with the tests."¹ The five tests referred to are those named above which were given before the school was opened. The estimates of the teachers were for intelligence and industry. A more detailed discussion of this question and also a fuller description of the entrance of these boys into the Speyer School may be had by consulting the study referred to above.

It is also of importance in connection with this study to point out that the boys were divided into groups for the purpose of instruction on the basis of their achievement in terms of their average rank in the eleven tests. When the average rank for each boy was determined groups of about twenty-five each were formed on the basis of achievement in the tests. At any time after this the teachers by their combined judgments could make any transfers they considered desirable so long as the groups were kept approximately the same in size.

Of the ninety-seven boys who were given the February 1916 tests seventy-five were in Speyer School in February, 1917, and seventy-two in June, 1917, when the collection of data for this investigation was finished. The distribution of these seventy-two boys among the different groups in June, 1916 and June, 1917 is shown in Table VII.

TABLE VII
DISTRIBUTION OF THE 72 BOYS AMONG GROUPS IN SCHOOL

<i>Groups</i>	1	2	3	4	5	6
June, 1916.....	13	11	12	11	14	11
June, 1917.....	13	11	15	7	16	10

This table shows that the seventy-two boys were quite uniformly distributed among the two hundred and therefore were not materially different from typical seventh grade boys.

After the February 1917 tests were given the boys were numbered from 1 to 75 according to the alphabetical arrangement of their names. These serial numbers are retained throughout

¹ Fretwell, E. K., *A Study in Educational Prognosis*, Teachers College, Columbia University, Contributions to Education, No. 99.

the investigation. Tables XXXIX to XLI in the Appendix contain the scores for the three testings. In February 1917 complete records for seventy-five individuals were secured. Three boys, Nos. 4, 22, and 37, were not present when the June 1917 tests were given. Because some of the statistical work had been done before the last testing was made the serial numbers were not changed to 1 to 72 but instead the three numbers noted above were dropped.

2. THE ADMINISTRATION AND SCORING OF THE TESTS

The standardized educational and psychological tests listed in Table VIII were used to secure the data for studying the educational attainments of the pupils who have been described above. The tests and the method of administering them will now be described briefly. References to full discussion of the tests by their authors are given for those not already familiar with these tests who may wish to make further study of them.

*Woody Arithmetic Scales*²

The Multiplication Scale, Series A, consists of thirty-nine problems scaled in degree of difficulty. The first problem is so easy that out of 943 seventh grade pupils tested by the author of the scale 936 solved it correctly, and the last one so difficult that of the same group only 186 solved it correctly. Multiplication Scale, Series B, is composed of twenty problems selected from Series A. It covers practically the same range of difficulty as does Series A.

The Division Scale, Series A, is made up of thirty-six problems, the first of which was solved by 822 out of 940 seventh grade pupils tested by the author of the scale, and the thirty-sixth by 123. Fifteen problems of Series A covering the entire range of difficulty compose Series B.

The time given was sufficient for practically all of the pupils to complete the tests. In accordance with the recommendation of the author "the standard for marking a problem correct was absolute accuracy, and, wherever possible, reduction to its lowest terms." One point was given for each correct answer. The score for the individual is the number of correct answers.

² Woody, Clifford, *Measurements of Some Achievements in Arithmetic*, Teachers College, Columbia University, Contributions to Education, No. 80.

*Hotz Algebra Scales*³

The Hotz First Year Algebra Scales were in the process of construction when the data for this study were secured. The Addition and Subtraction Scale is made up of twenty-four problems scaled in degree of difficulty from easy to difficult problems. The Multiplication and Division Scale is composed of twenty-three problems. It is built on the same principle.

*Trabue Completion-Test Language Scales*⁴

These tests are composed of mutilated sentences which the subject is to complete by filling in the words which make the "most sensible statement." Scales B, C, D, and E consist of ten sentences each, scaled so that they range in P.E. units of value from about 1 to between 10.5 and 11. The intervals between sentences are nearly equal. Scales J and K have seven sentences each, ranging in value from a little more than 4 to about 12.5; and L and M have eight sentences each, ranging from almost 7 to a little above 11 P.E. units of difficulty.

Seven minutes were given for completion of the sentences of each scale. In this amount of time all the subjects apparently had opportunity to show their maximum ability in such work for in most cases more sentences were attempted than were correctly done. The method of scoring was that suggested by the author of the scales. In cases where the lists given in his guide for scoring did not cover the answer in question the standard decided upon was recorded and used in any similar instances. This made for uniformity in scoring. Two points were given for each sentence completed correctly and one point for "each sentence completed with only a slight imperfection."

*Scale Alpha 2. For Measuring the Understanding of Sentences*⁵

Part II of this scale was used. Scale Alpha 2 is "an improved and extended form" of "a provisional scale Alpha for measuring ability in paragraph reading." Part II begins with

³ Hotz, Henry G., *First Year Algebra Scales*, Teachers College, Columbia University, Contributions to Education, No. 90.

⁴ Trabue, M. R., *Completion-Test Language Scales*, Teachers College, Columbia University, Contributions to Education, No. 77.

⁵ Thorndike, E. L., "Measurement of Achievement in Reading," *Teachers College Record*, Vol. XV, No. 4. "An Improved Scale for Measuring Ability in Reading," *Teachers College Record*, Vol. XVI, No. 5 and Vol. XVII, No. 1.

difficulty 7 and extends through difficulties 8, 8½, and 9. There are ten paragraphs in all, concerning the meaning of which twenty-four questions are asked. The subject's achievement in the test is determined by his answers to the questions asked on each paragraph.

The selections from Beta and from S are similar to the paragraphs in Alpha 2. Because Alpha 2 had been used twice it was considered best not to repeat it again. Therefore three paragraphs were selected from Scale Beta, and the one paragraph of S of the longer reading scale was added. Twenty-nine questions are asked concerning the meaning of these paragraphs.

In scoring the tests answers were divided into three classes, —correct, slightly incorrect, and wrong, for which 2, 1, and 0 points respectively were given. The total number of points is the score given the individual. The time was sufficient for all but the very slowest to do as much as they could with the test. As an aid to uniformity in scoring record was made of types of answers concerning which there was question as to their classification. This was used to supplement the list given by the author of the scale.

Visual Vocabulary ⁶

The Visual Vocabulary tests consist of lists of words which are to be classified accordingly as they mean a flower, an animal, a boy's name, a game, a book, something about time, something good to be or do, or something bad to be or do. The classification of the word is indicated by writing a designated letter or word under it.

The Thorndike Reading Scale A was given in February, 1916. It consists of forty-three words arranged by groups of five in ascending degrees of difficulty. The last group has only three words. The test given in February, 1917 was made up of one hundred and seventy words in fourteen groups selected from the Thorndike Scale A 2 plus four groups selected from its provisional extension. The groups begin with step $6\frac{1}{2}x$ and extend through step $12\frac{1}{2}$.

The Thorndike Reading Scale B, *y* series, was given in June, 1917. It consists of one hundred and twenty words arranged

⁶ Thorndike, E. L., "Measurement of Achievement in Reading," *Teachers College Record*, Vol. XV, No. 4, and Vol. XVII, No. 5.

in groups of ten. It is built on the same principle as the other two tests, using however a different list of meanings to determine the classification.

*Composition*⁷

The subjects for the test in composition were: for February, 1916, How I Would Spend Twenty Dollars; for February, 1917, What I Should Like to do Next Saturday; for June, 1917, How I Should Like to Spend My Vacation, or A Narrow Escape. These were all rated by the Hillegas Scale for the Measurement of Quality in English Composition. The first set of compositions was rated by from four to eight experienced judges. The average of their marks was taken as the score for the composition. The second set was rated by four experienced judges and the third set by three of the four who rated the second set. Here also the ratings were averaged to determine the score for the composition.

The time allowed for writing the composition was thirty minutes for the first two sets and fifty minutes for the third set.

*Spelling*⁸

The Ayres Measuring Scale for Ability in Spelling was used. The first time the tests were given fifty words were selected from the Q list. The Q list is rated as of a difficulty such that the average score of a seventh grade class should be 92 per cent. In the second testing fifty words selected from lists U, V, W, and X were given. The last time fifty words from lists T to Z inclusive were used.

The words were pronounced by the regular teacher. Each word was pronounced twice and a third time if asked for. One point was given for each word spelled correctly. The teachers did not score the papers.

*Opposites Tests*⁹

The Opposites Tests consist of twenty words each. The purpose of the test is to determine the number of words having

⁷ Hillegas, Milo B., "A Scale for the Measurement of Quality in English Composition by Young People," *Teachers College Record*, Vol. XIII, No. 4.

⁸ Ayres, Leonard P., *A Measuring Scale for Ability in Spelling*, Division of Education, Russell Sage Foundation, Bulletin E 139.

a meaning opposite the words of the list, which can be written in a given length of time. The "north-south" and the "long-short" lists are of equal difficulty. These were used for the first two testings. The time allowed was seventy-two seconds. The "high-low" list is made up of the easiest words of the other two and consequently the time was reduced. Forty seconds were allowed when it was given in June, 1917.

The responses were classed as either right or wrong. One point was given for each correct response.

*Mixed Relations Test*⁹

In the Mixed Relations Tests a pair of words is given to indicate the relation desired in each response to a third word. There are twenty such series in each test. Before the test began a sample was exhibited and the explanation made that after the third word of each series a fourth word was to be written which would have the same relation to the third word that the second had to the first. In the first two testings one hundred and twelve seconds were allowed, but the third time this was reduced to ninety seconds. The responses were considered either right or wrong, one point being given for each one right.

*Easy Directions Test*⁹

In the Easy Directions Tests the subject is directed to make a definite response such as: Cross out the smallest dot . . . , or Cross out the g in tiger. The two tests are of approximately equal difficulty. The "smallest dot" test was given in February, 1916, and the "g in tiger" test both times the tests were repeated. One point was given for each correct response. The time allowed was eighty-two seconds for the first two testings and eighty for the third.

*Hard Directions Test*⁹

The Hard Directions Test is similar to the Easy Directions except that "the object here is to complicate the directions somewhat, by calling for conditional and alternative responses, etc." The first two or three directions are easy enough to in-

⁹ Woodworth, R. S., and Wells, Frederic Lyman, "Association Tests," *Psychological Monographs*, Vol. XIII, No. 5.

sure a proper start on the test and the rest are more complicated. Because of the "conditional and alternative" responses the scoring is somewhat complicated. A standard of twenty-two possibilities for mistakes was decided upon and used consistently in scoring. From "twenty-two" one was deducted for each wrong response. The time allowed was two minutes.

All of the tests were either scored or their scorings checked by one or the other of the two persons chiefly interested in the prosecution of this study,—except the Composition tests, the scores of which as has already been stated are averages, the Algebra tests, and twenty-eight of the seven hundred and ninety-two papers of the February 1916 testing. In scoring the papers and copying the scores extreme care was taken to avoid chance mistakes. This increased the amount of time consumed, but greater accuracy in scoring is needed for individual results than for group results.

It is very essential to the purposes of this investigation that the scoring be uniform and that as fine discriminations as possible be made because the achievement of the individual in specific tests is the problem for study. An error in scoring which affects the group standing only slightly when carried over to the individual, although the same in absolute value, has relatively a much greater significance in the case of the individual.

3. SPECIAL TESTING

After the third testing of the entire group was completed in June, 1917, a special testing of certain boys was made to compare their reactions under conditions of more detailed control. In Section VI the results obtained from the special testing are analyzed and compared with the results from the original testings. All three Spelling tests were repeated with three boys. The words were pronounced by the writer. Each word was pronounced twice and a third time if necessary. The "long-short" and "high-low" Opposites tests were repeated with four boys. The time for each was forty seconds. Five boys were given both Mixed Relations tests. Time: ninety seconds for each. Both Easy Directions tests were repeated with four boys, the time allowed being eighty seconds. These special tests were given in the office at the Speyer School. Not more than three

boys were tested at any one time. Since in these cases a low score had been made in one or more of the original tests it was suggested that probably this was caused by some disturbance or that the boy was not feeling well on the day of the test; further, that probably he could do better and that an opportunity was then going to be given. The same explanation that was given at the original testing was made.

TABLE VIII

THE TESTS AND THE TIMES AT WHICH THEY WERE GIVEN

<i>February, 1916</i>	<i>February, 1917</i>	<i>June, 1917</i>
Woody Multiplication Series A	Woody Multiplication Series B	Hotz Algebra Add. and Subt.
Woody Division Series A	Woody Division Series B	Hotz Algebra Mult. and Div.
Trabue Completion Scale B	Trabue Completion Scale J	Trabue Completion Scale L
Trabue Completion Scale C	Trabue Completion Scale K	Trabue Completion Scale M
Reading Alpha 2 Part II	Reading Alpha 2 Part II	Reading Selections from Beta and from S
Visual Vocabulary Reading Scale A	Visual Vocabulary Selection from Scale A 2	Visual Vocabulary Scale B y series
Composition How I Would Spend Twenty Dollars	Composition What I Should Like to do Next Saturday	Composition How I Should Like to Spend My Vacation
Spelling 50 words from Ayres Q List	Spelling 50 words from Lists U to X	Spelling 50 words from Lists T to Z
Opposites North—South	Opposites Long—Short	Opposites High—Low
Mixed Relations Good—Bad	Mixed Relations Eye—See	Mixed Relations Good—Bad
Easy Directions Smallest dot	Easy Directions G in tiger	Easy Directions G in tiger
<i>June, 1916</i>	Hard Directions *	Hard Directions *
Trabue Completion * Scale D		
Trabue Completion * Scale E		

* These tests were used for a slightly different purpose from that of the eleven tests above.

IV

STATISTICAL TREATMENT

1. TRANSMUTATION AND DISTRIBUTION OF SCORES

When all the papers had been scored the first step was to record the scores of the seventy-two boys in such manner that the score of every boy in each test could be identified. Tables XXXIX to XLI in the Appendix contain these results. A distribution table of the original scores was then made for each of the thirty-seven tests. The semi-interquartile-range (Q) of each of these

distributions was found by using the formula: $\frac{\frac{3}{4}n - \frac{1}{4}n}{2}$

These Q's are given in Table IX. The reason for using the Q instead of the S.D., which was used in the preliminary investigation, is discussed under topic two of this section.

In order that a part of the statistical work could be done before the last tests were given and scored in June 1917, the scores of the group of seventy-five pupils who took the eleven tests in February 1916 and February 1917 were used for the distributions and transmutations. In June 1917 three of these seventy-five pupils were not present when the tests were given. Their scores in the two previous testings were dropped from further consideration in this study. This produced practically no change in the Q's from what they would have been if only the seventy-two pupils' records had been used to find the Q's, especially since of the three records missing one was in the first tertile and two in the second tertile in February 1916, and one in each tertile in February 1917.

Using the Q's shown in Table IX the intervals of each distribution according to the original scores were transmuted into intervals according to their value in terms of the Q of the original distribution. The frequencies in these intervals, grouped in intervals of one Q, are shown in Table X. They are graphically represented by Figs. 3a, b, c, to 13a, b, c.

TABLE IX
THE SEMI-INTERQUARTILE-RANGE (Q) OF THE DISTRIBUTION OF THE
ORIGINAL SCORES FOR EACH TEST

Tests	Feb. 1916	Feb. 1917	June 1917	June 1916
Woody Multiplication	2.13	1.21		
Woody Division	2.62	.67		
Hotz Alg. Add. and Subt.....			2.06	
Hotz Alg. Mult. and Div.....			2.67	
Trabue B, J, L, D.....	1.71	1.22	1.10	1.36
Trabue C, K, M, E.....	1.32	1.45	2.03	1.58
Reading Tests	4.26	4.59	3.13	
Visual Vocabulary	3.77	13.88	4.27	
Composition	5.69	4.17	5.14	
Spelling	1.60	3.94	3.13	
Opposites	1.89	.47	1.54	
Mixed Relations	4.58	2.95	2.93	
Easy Directions	2.44	1.49	.76	
Hard Directions		2.46	1.57	

TABLE X
DISTRIBUTION OF SCORES TRANSMUTED INTO MULTIPLES OF Q ABOVE AND
BELOW THE MEDIAN OF THE ORIGINAL DISTRIBUTION IN EACH TEST

February 1916									
Value in Q	Woody Mult.	Woody Div.	Trabue B	Trabue C	Reading Alpha 2	Visual Vocabulary	Composition	Spelling	Opposites
+4.0 to +4.9				2			1		
+3.0 to +3.9					1	3	1		
+2.0 to +2.9	4		8	11	7	5	8		
+1.0 to +1.9	14	15	16	5	14	11	10	7	14
.0 to +.9	18	21	12	18	14	17	18	29	22
.0 to -.9	16	14	21	23	22	18	19	12	11
-1.0 to -1.9	9	13	9	11	9	7	13	12	14
-2.0 to -2.9	7	7	3	1	4	3	4	12	7
-3.0 to -3.9	2	1	2		1	3		3	4
-4.0 to -4.9			1	1				3	
-5.0 to	2	1						3	

February 1917									
Value in Q	Woody Mult.	Woody Div.	Trabue J	Trabue K	Reading Alpha 2	Visual Vocabulary	Composition	Spelling	Opposites
+4.0 to +4.9			2	2		2	3		
+3.0 to +3.9	1					2	7		
+2.0 to +2.9	4		5	6	2	2	9		
+1.0 to +1.9	13	18	5	17	18	16	17	14	
.0 to +.9	18	18	24	11	16	16	17		11
.0 to -.9	12	13	23	28	18	21	22	38	25
-1.0 to -1.9	15	15	9	5	8	11	16	16	8
-2.0 to -2.9	6	5	4	2	3	2	2	7	7
-3.0 to -3.9	2			1	3	2		2	5
-4.0 to -4.9		1			1			2	2
-5.0 to	1	2						2	7

June 1917											
Value in Q	Algebra Add. Subt.	Algebra Mult. Div.	Trabue L	Trabue M	Reading	Visual Vocabulary	Composition	Spelling	Opposites	Mixed Relations	Easy Directions
+4.0 to +4.9			3								
+3.0 to +3.9	4		1	1			1				
+2.0 to +2.9	9	6	2	3		1	9				
+1.0 to +1.9	8	15	19	7	12	12	12	16	19	10	30
.0 to +.9	15	15	11	25	24	23	14	20	17	26	6
-.0 to -.9	26	18	22	12	13	15	21	13	19	15	17
-1.0 to -1.9	7	14	9	19	10	10	13	12	8	10	9
-2.0 to -2.9	2	4	4	4	7	5	2	2	6	5	5
-3.0 to -3.9	1		1	1		1	2	4	2	3	
-4.0 to -4.9					3	1		4	1	1	2
-5.0 to —					3	4		1		2	3

2. THE USE OF AVERAGES AND VARIABILITIES

Two methods of comparing the scores of an individual in all the tests were discussed in Section II. Under this topic of this section the reason for a slight variation of the method selected will be given, and excerpts from studies in which the method has previously been used will be quoted.

Because one purpose of this study is to discover the extremely variant scores,—those due to causes that prevent the individual from making a normal or characteristic reaction, and also those due possibly to unusual ability or the lack of ability—a method that would tend to cover up these sport scores should not be used, but rather a method which retains their variability in relative proportions should be used. The standard deviation as a measure of variability gives more weight to the extreme items than to those nearer the central tendency. Therefore it would seem that a measure of variability which avoids such weighting should be chosen. The Q was chosen because with it the range of the items beyond it does not affect the measure of variability,—the number of items beyond a given point being the influencing factor.

A minor reason for using the Q rather than the S.D. is found in the matter of mathematical inaccuracies introduced by the lopping off of fractions. The measures of the group are carried to two decimal places, the second being an approximation determined by the size of the third, and the measures of the individual are carried to one decimal place only with the same method

of approximation. In cases of five-tenths or more the preceding figure is increased by one; in cases of less than five-tenths it is not changed. The S.D. being 1.4825 Q means that it would carry with it a relatively greater mathematical approximation in each case.

The following excerpts, one from a study by Naomi Norsworthy and the other from a study by R. S. Woodworth, describe and discuss the method more fully.

We now have two series of grades in the same measurement, one set from mentally defective children and the other from ordinary school children. The usual method of comparing such results is to compare the records of one set of individuals with the central tendency of those of the others of the same age and sex. But in this case there were not enough defectives of any age to make the results gained from such treatment of any value, consequently a different method has to be adopted. The method used in dealing with the majority of the measurements was one which enabled me to compare the records of all of the defectives with those of all the ordinary children without restriction as to age or sex. Another very decided advantage is the fact that the units of grading are identical throughout all the measurements, as will be evident from the following description.

The difference between the record of each defective in any test and the median for an ordinary child of the same age and sex was found. This difference was then transmuted into positive or negative multiples of the probable error as the case required. . . . By thus transmuting the difference between the grading received by defectives and ordinary children respectively in every test into multiples of the probable error of the appropriate age and sex I can compare the records of the 150 defectives tested with the 500 or 600 ordinary children, just as if I had 150 idiots and 600 school children all of the same age and sex. Not only by this method can I consider all my cases together, but each test is, so far as is possible comparable with every other, irrespective of whether the trait examined is physical or mental. This, so far as I know, has not yet been done. . . . This method, then, provides a measure by which we can tell not only how far the idiots are below school children in the various traits tested, but how much farther below they are in one mental trait than in another and whether they are equally deficient in physical and mental traits.¹

What is needed is a method of combining results which shall preserve all the refinement of the original measurements. Such a method exists, and is certainly familiar to statisticians; but it seems to be overlooked in many cases where it would prove of value.

Here follows a discussion of averages and variabilities.

There is a way of eliminating both of the troublesome quantities—both the absolute value of the average and the absolute measure of variability. Let the average in each case be counted as 0, i.e., let the individual's standing be expressed as a deviation above or below the average; and further, let the measure of variability be taken as the unit deviation, and

¹ Norsworthy, Naomi, *The Psychology of Mentally Deficient Children*, Columbia University Contributions to Philosophy and Psychology, Vol. XV, No. 2, p. 50.

all deviations be expressed as fractions or multiples of this unit. (For the measure of variability, either the average deviation, or the mean square deviation, or the quartile, etc., may be chosen.) What this method does is to assign each individual's position in the distribution of the group: he stands, namely, above or below the group average, and so and so much above or below as compared with the average variation of the group.

No assumption is made by this method as to the ratio between the variability and the group average; for the average is taken as 0 and the variability as 1, independently the one of the other. The only assumptions underlying the method are those involved in every use of averages and variabilities, namely, that the average means the same thing in respect to one distribution as in respect to another, and, likewise, that the measure of variability means the same thing in respect to the different distributions. Both of the assumptions are correct if the distributions are of the "normal" type, or if all the distributions belong to any one type. Were one distribution normal, another markedly skew, and a third distinctly bimodal, neither the average nor the average deviation would mean quite the same thing in respect to the three, and the method would be illegitimate; but in such a case it is doubtful if the distributions ought properly to be combined at all. Mental tests usually give group distributions not very different from the "normal" though tending on the whole to be somewhat skew in such a way that more individuals lie on the good side than on the bad side. The distributions for different tests do not differ much in shape, and no considerable error can be introduced by placing the average always equal to 0 and the average deviation (or mean square deviation, etc.) always equal to 1.²

Although the method is fully described by the two quotations just given, a concrete illustration from this investigation may serve as an aid in projecting the method to this study. In Fig. 2 are represented graphically the records of the same three individuals whose scores according to the classification by rank are shown in Fig. 1. The reduction of the number of subjects from ninety-seven in the preliminary investigation to seventy-two in the study proper changed the serial numbers so that Case 60 here is Individual 45; Case 65, Ind. 49; and Case 92, Ind. 71. The reduction of the number in the group makes it unfair to compare directly the range and placement of scores in values of *Q* with their range and placement by rank. Further comparison in this respect will be made in Section VI. Fig. 2 is introduced here to illustrate the method.

To compare these individuals in their standing in the same test, in terms of achievement in that test, their scores should be read by the scale at the right of Fig. 2. In spelling (*S*) for example, Ind. 45 is .1*Q* above the median of the group (*Med. gr.*), likewise, Ind. 49 is .1*Q* above the median of the group. They both received the same original score in spelling which

² Woodworth, R. S., "Combining the Results of Several Tests; A Study in Statistical Method," *The Psychological Review*, Vol. XIX, pp. 97-101.

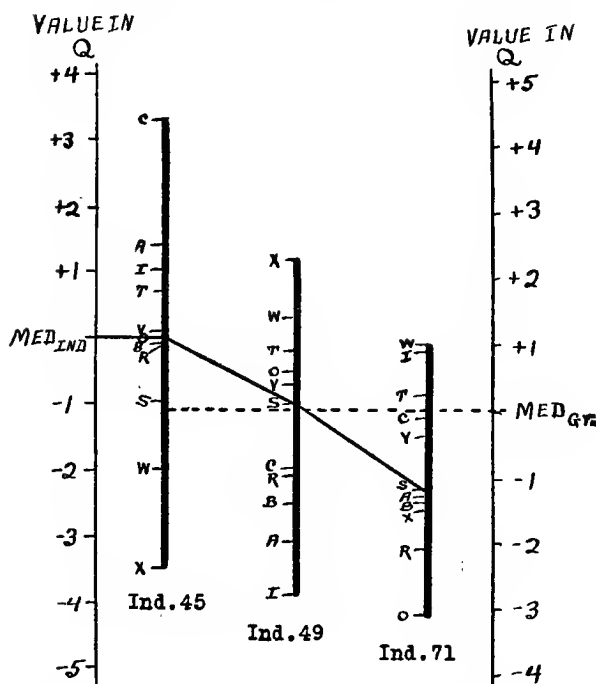


Fig. 2. Distribution of the Scores, in Values of Q , of Three Individuals. February 1916 Tests. The letters signify tests as follows:

X—Woody Multiplication
W—Woody Division
B—Trabue B
C—Trabue C
A—Reading Alpha 2
V—Visual Vocabulary

T—Composition
S—Spelling
O—Opposites
R—Mixed Relations
I—Easy Directions

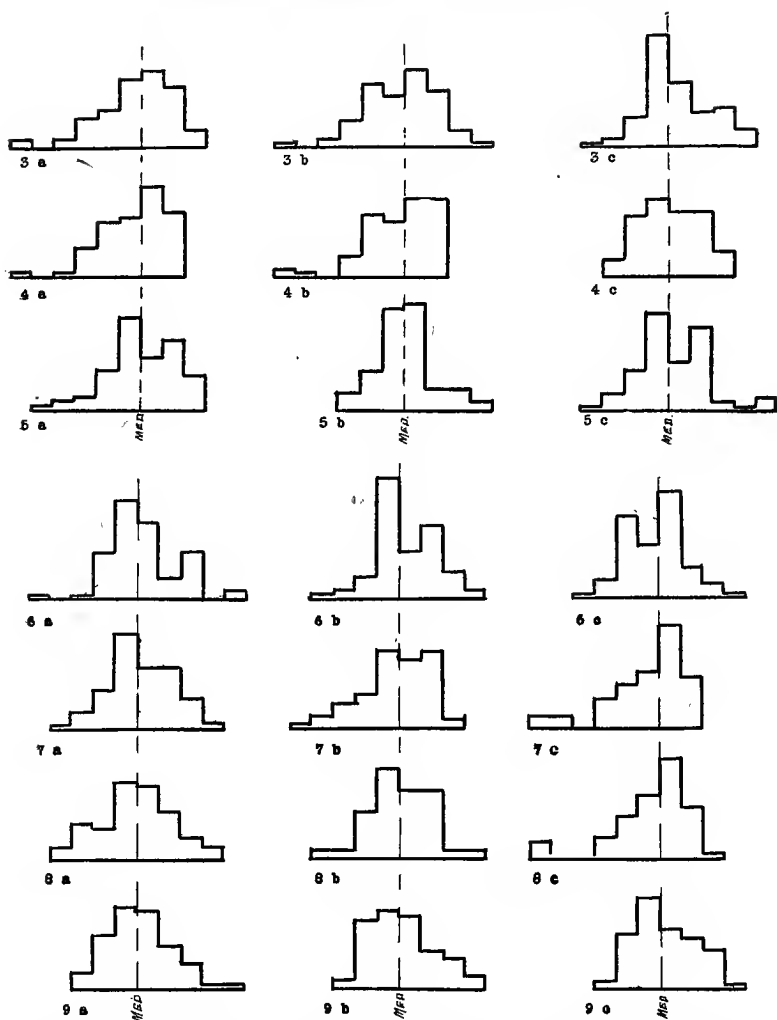
was 48. Ind. 71 is 1.2 Q below the median of the group in spelling. His score was 46. The achievement in this spelling test is then the same for Individuals 45 and 49, and 1.3 Q less for Ind. 71. The median achievement in one test is rated equal to the median achievement in another, and likewise, the Q deviation in one test is equated with that of any other.

This explanation holds also in comparing the individual's achievement in different tests and the achievement of different individuals in different tests. For example, in relation to the median achievement of the group in all the tests Ind. 45 achieved the same in Trabue B and Mixed Relations as did Ind. 71 in Woody Division. All three of these scores are 1.0 Q above the

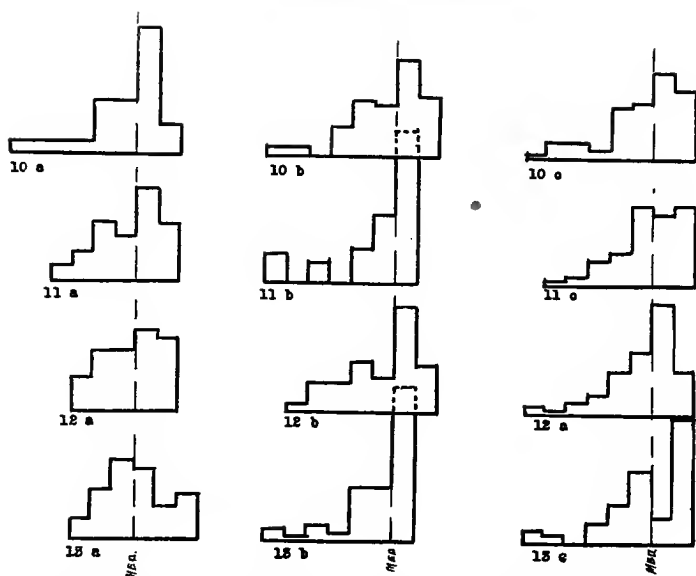
median of the group. The achievement of one individual in different tests is simply stated in multiples of Q above and below the median of the group, and therefore, compared with the median of the group in all the tests by reading from the scale at the right of the figure.

To compare these individuals in their standing in the same test, in terms of their own median achievement, their scores should be read by the scale at the left of Fig. 2. The median line of this scale begins with the median score of the individual ranking highest as judged by median attainment, and connects the median scores of the three individuals. Using spelling again the figure shows that in comparison with their other scores Inds. 49 and 71 are alike in achievement in spelling,—spelling represents their median achievement—while the spelling achievement of Ind. 45 is 1.0 Q below his median achievement. This shows that the original scores cannot be taken at face value in diagnosing individual cases. This answers the question in Section I concerning the meaning that the same scores in a single test may have in connection with the achievement of different individuals. Two scores of the same test having the same value by the original scoring may have very different meaning in the two distributions of the individuals' scores; one may be comparatively high among the achievements of one individual, while the other may be comparatively low. Any two scores can be related to the median achievement of the respective individuals by referring them, following the line marked out by the median of the individuals, to the scale at the left of the figure.

In the excerpt above Woodworth points out that the validity of not only this but of any method of comparing or equating scores depends to a large extent upon the similarity among the forms of distribution of the measures. Figs. 3a, b, c to 13a, b, c show graphically the distributions of scores in values of Q in thirty-three of the tests. The distributions show no great dissimilarity except in the cases of 11b, Opposites, February 1917; 13b, Easy Directions, February 1917; and 13c, Easy Directions, June 1917, which are rather markedly skewed. The other distributions show skewness upward and downward to varying degrees, but are similar enough in general to be compared with but little loss in accuracy from this cause.



The effect of the skewness of the original distributions mentioned above upon the Q is illustrated by Fig. 14. The piling up of scores reduces the extent of the semi-interquartile-range measured by intervals of the original distribution of the scores of the group. The reduction is more pronounced in the range above the median than in the range below the median. In this form of distribution the Q is 87.4 per cent of the Q of the nor-



Figs. 3a, b, c to 13a, b, c. Distribution of Scores in Each Test Transmuted into Multiples of Q Above and Below the Median of the Original Distribution.

Feb. 1916			Feb. 1917			June 1917		
Fig. 3a.	Woody Mult.		Fig. 3b.	Woody Mult.		Fig. 3c.	Alg., Add. Subt.	
" 4a.	Woody Div.		" 4b.	Woody Div.		" 4c.	Alg., Mult. Div.	
" 5a.	Trabue B		" 5b.	Trabue J		" 5c.	Trabue L	
" 6a.	Trabue C		" 6b.	Trabue K		" 6c.	Trabue M	
" 7a.	Read. Alpha 2		" 7b.	Read. Alpha 2		" 7c.	Reading	
" 8a.	Visual Vocab.		" 8b.	Visual Vocab.		" 8c.	Visual Vocab.	
" 9a.	Composition		" 9b.	Composition		" 9c.	Composition	
" 10a.	Spelling		" 10b.	Spelling		" 10c.	Spelling	
" 11a.	Opposites		" 11b.	Opposites		" 11c.	Opposites	
" 12a.	Mixed Relat.		" 12b.	Mixed Relat.		" 12c.	Mixed Relat.	
" 13a.	Easy Direct.		" 13b.	Easy Direct.		" 13c.	Easy Direct.	

mal distribution.³ The reduction in the value of the Q increases the minus value of the low scores expressed in multiples of Q. It places them, when compared with scores of tests having a normal distribution, at a greater distance from the median than they would be if the test had been of such nature that the higher scores had been spread out approximating more closely the form of the normal distribution. Moreover, the high scores which would have been higher if the test had permitted are covered up by this form of distribution. In Fig. 14 a score 3Q below the

³ Thorndike, E. L., *Mental and Social Measurements*, p. 73. Surface of Frequency of Form C. (The Q is derived from the values of σ and the surfaces A and C are smoothed.)

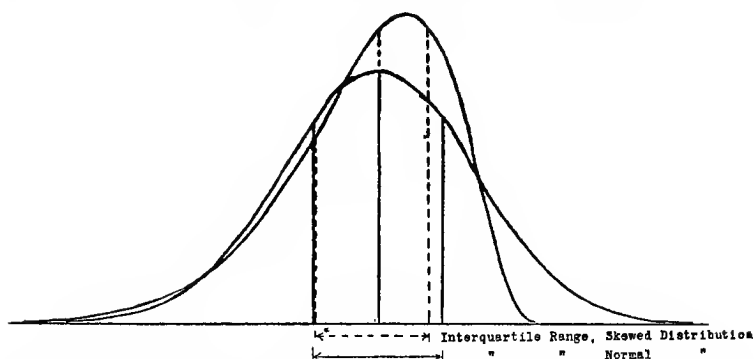


Fig. 14. Showing the Effect of Two Different Forms of Distribution upon the Q .

median of the normal distribution would be 3.43 Q below the median in the skewed distribution.

3. REDISTRIBUTION OF SCORES

After the intervals of the original distributions were transmuted into multiples of Q the next step was to transmute each score into a multiple of the Q of its distribution. A table containing such transmutations of the 2664⁴ scores grouped vertically for the tests and horizontally for the individuals was made. From a part of this table,—that part showing the transmuted scores of eleven of the June 1917 tests—the chart shown in Fig. 15 was constructed.

As already suggested under Topic 2, achievement by any two scores of one test or different tests can be compared directly by this method. The chart of Fig. 15 facilitates such comparison of the scores of the June 1917 tests. The scales for the graph are the same as those of Fig. 2,—the one at the right gives values of Q above and below the median of the group, and the one at the left, values above and below the individual medians.

For use in the calculations made, charts similar to this were

⁴ This would be the number if every boy had taken every test. Scores are lacking for pupil No. 40 in Composition and Spelling, June, 1917, and for pupil No. 27 in Woody Division, February, 1917. In these cases the median score of the group was supplied. Four scores are lacking in Trabue D and E each. These were not supplied because these two tests were not used in the group of eleven.

Serial Number of Individual

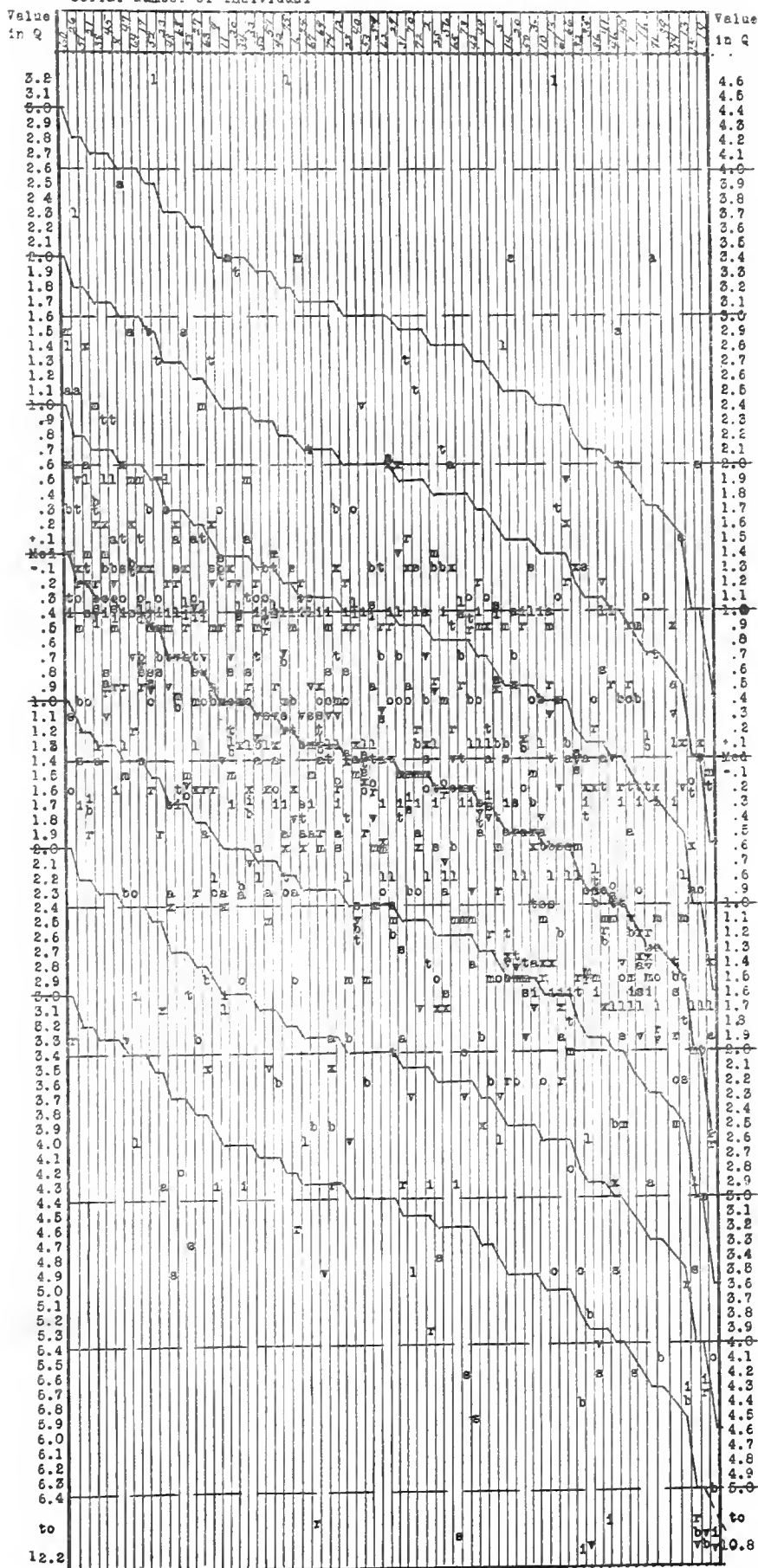


Fig 15. Chart Showing the Value of Q in Each Score of the June 1917 Tests. The letters signify tests as follows:

a—Algebra, Add. Subt.	b—Reading	o—Opposites
x—Algebra, Mult. Div.	v—Visual Vocabulary	r—Mixed Relations
l—Trabue L	t—Composition	i—Easy Directions
m—Trabue M	s—Spelling	

The scale at the left of the chart is for values above and below the median of the individuals. The scale at the right of the chart is for values above and below the median of the group.

constructed for each of the other two testings using the eleven tests, the eight Trabue tests combined, the six mathematics tests combined, the five directions tests combined, and the three reading tests combined. In these charts each score of each individual is identified by a letter so that any two of them can be found and compared in respect to achievement in values from the median of the group or in relation to the other achievements of the individual.

Up to this point the experimental material used in this study and the method of treating this material statistically have been considered. The subjects and the tests have been described. The subjects represent very closely typical seventh grade ability. The tests used were not devised for this special study but are tests which have been carefully standardized and used extensively in other investigations. A slight variation from the method of equating scores decided upon in the preliminary investigation has been discussed. The *Q* rather than the *S.D.* is used as the measure of variability because the extreme scores,—those which probably do not represent the individual's normal reaction—have no greater effect upon the *Q* than do other scores; while they do have a greater effect upon the *S.D.* than other scores nearer the central tendency have. It has been shown that in a given test two scores having the same value do not necessarily have the same meaning for the two individuals; but that the meaning of each score must be interpreted by comparison with the other achievements of the individual. For example, one of two scores equal in value may be very low for one individual in comparison with his other achievements while the same score for another individual may be equal to or above his median achievement. The next section will deal with the results found in connection with individual variability as compared with group variability.

V

INDIVIDUAL VARIABILITY COMPARED WITH GROUP VARIABILITY

1. THE AMOUNT OF INDIVIDUAL VARIABILITY

What is the amount of the individual's variability among the different tests? Stated more specifically do the scores of some individuals tend to be high, do the scores of others tend to be near the average of the group, and do the scores of still others tend to be low? Or are the scores of most individuals so spread out that there is no well defined mode? The answers to these questions involve other questions, namely: By what standard shall the individual's variability be measured, and by what method can the measurement be made?

The unit of measurement already described will be used. The *Q* of the group will be taken as the unit or standard. The median achievement will be taken as the starting point and variability will be measured in terms of the amount of the deviation in either one or both directions from the median. What then is the amount of variability of the group? It is the standard or unit, one *Q*. Having now related the standard to the problem under consideration the question can be asked in more specific terms, namely: What per cent of the variability of the group in each test is the variability of the individual in all the tests,—that is, what per cent of the *Q* taken as the standard is the variability of the individual?

With the scores transmuted into values of *Q* and redistributed in charts such as that shown in Fig. 15, the range between the two extreme scores, the range between the median score and the last score above the median, the range between the median score and the last score below the median, and the range between the third score above and the third score below the median were found for each individual. Using a scale these were read directly from the charts. The averages of these ranges for the group and for different divisions of the group are shown in Table XIII. The last of the ranges enumerated above, the range be-

Individual Variability Compared with Group Variability 31

tween the third score above and the third score below the median, is an approximation of the interquartile range. It covers a distance of 3 intervals on each side of the median whereas the interquartile range covers a distance of only 2.75 intervals on each side. It was used as a matter of economy in calculation. This distance was readily determined whereas the distance of 2.75 intervals would have necessitated interpolation for the value in every case. The following correction was made for the approximation of the interquartile range so that the results would be comparable with the Q of the original distributions.

By the method used the approximation of the interquartile range covers a distance of 6 intervals, 3 intervals on each side of the median. The interquartile range covers a distance of 5.5 intervals, 2.75 intervals on each side of the median. The extent, in terms of the Q of the original distributions, of $3/5.5$ of the measures on each side of the median was found. The extent of $2.75/5.5$ of the measures on each side of the median is desired. $3/5.5$ of 50 per cent = .2727. Using a table¹ of values of x/Q of the normal probability integral it is found that 27.27 per cent of the surface in each direction from the median corresponds to a distance of approximately 1.11 Q on the base line. Hence the values found are 111 per cent of the values desired. Dividing the values given in Table XIII (E) by 2 and making this correction we have the values of the semi-interquartile range or Q which are given in Table XI.

TABLE XI
AVERAGE OF THE INDIVIDUAL SEMI-INTERQUARTILE-RANGES IN THE
ELEVEN TESTS

The table reads as follows: In February, 1916, the average of the Q's of the first quartile was 80 per cent of the Q taken as the standard, etc.

	Quartile				Tertile			Total	Corrected Total
	I	II	III	IV	I	II	III		
February 1916..	.80	.90	.91	.99	.80	.93	.96	.90	.81
February 1917..	.71	.86	.87	1.17	.75	.82	1.14	.90	.81
June 1917.....	.79	.75	.94	1.29	.76	.88	1.19	.94	.85
Average	.77	.84	.91	1.15	.77	.87	1.10	.91	.82
Corrected									
Average.....	.69	.76	.82	1.04	.69	.78	.99	.82	

¹ Thorndike, E. L., *Mental and Social Measurements*, p. 220.

Inspection of this table gives an answer to the question raised above, namely: What is the amount of the individual's variability among the different tests? Measured in terms of *Q* the average individual variability is 82 per cent of the variability of the group. That is, the average semi-interquartile-range of the individual is 82 per cent of the average semi-interquartile-range of the group. The average range in ranks of the ninety-seven pupils studied in the preliminary investigation and given in Section II was found to be 82 per cent of the total range possible. These two figures supplementing each other as they do, are convincing evidence of the large amount of variability among the achievements of these pupils in the different tests. This variability is evidence of the unreliability of one test or a small number of tests used for the purpose of educational prognosis.

The table shows further that the individual variability is greater in the third testing than in either of the first two, but not enough greater to be of significance. The difference in variability among the different divisions of the group as ranked by median achievement is consistent enough and large enough to be significant. When grouped either in quartiles or tertiles the lower ranking pupils are found to be more variable in their achievements. The corrected averages show that the variability of the fourth quartile is 50 per cent greater than the variability of the first; and that the variability of the third tertile is 43 per cent greater than that of the first. The fourth quartile exceeds the standard adopted and the third tertile almost equals it.

TABLE XII
DISTRIBUTION OF THE INDIVIDUAL SEMI-INTERQUARTILE-RANGES
(APPROXIMATION) IN THE ELEVEN TESTS

<i>Value in Q</i>	<i>Feb.</i> 1916	<i>Feb.</i> 1917	<i>June</i> 1917
2.0 to 2.4			2
1.5 to 1.9	3	5	4
1.0 to 1.4	22	20	25
.5 to .9	43	41	34
.0 to .4	4	6	7

Table XII gives the distribution of the individual semi-interquartile-ranges (approximation) for the three testings. It shows a slight increase in individual variability the longer the pupils remain in school. It means that there is a slightly greater range in the achievements of the individual pupils in the dif-

Individual Variability Compared with Group Variability 33

ferent tests. However, this difference is not great enough to base any conclusions upon it.

The amount of individual variability can be further measured by finding the total range, the range above the median, and the range below the median for each individual in his different achievements. The results are given in Table XIII.

TABLE XIII

AVERAGES IN CONNECTION WITH INDIVIDUAL RANGES IN SCORES
TRANSMUTED INTO MULTIPLES OF Q

The table reads as follows: In February, 1916, the average total range of the pupils in the quartile ranking highest was 4.14 Q, etc.

(A) AVERAGE TOTAL RANGE IN THE ELEVEN TESTS

	Quartile				Tertile			Total
	I	II	III	IV	I	II	III	
February 1916	4.14	4.14	5.39	5.04	4.03	5.15	4.85	4.68
February 1917	4.71	4.69	4.43	5.68	4.65	4.45	5.52	4.88
June 1917	4.21	3.97	4.59	6.42	4.05	4.55	5.79	4.87
Average	4.35	4.27	4.80	5.71	4.24	4.72	5.39	4.78

(B) AVERAGE RANGE ABOVE INDIVIDUAL MEDIANS IN THE ELEVEN TESTS

	Quartile				Tertile			Total
	I	II	III	IV	I	II	III	
February 1916	1.70	2.07	2.00	2.14	1.72	2.16	2.05	1.98
February 1917	1.48	1.78	1.75	2.18	1.49	1.89	2.02	1.80
June 1917	1.73	1.79	1.98	2.65	1.73	1.83	2.55	2.04
Average	1.64	1.88	1.91	2.32	1.65	1.96	2.21	1.94

(C) AVERAGE RANGE BELOW INDIVIDUAL MEDIANS IN THE ELEVEN TESTS

	Quartile				Tertile			Total
	I	II	III	IV	I	II	III	
February 1916	2.44	2.07	3.39	2.90	2.31	2.99	2.80	2.70
February 1917	3.22	2.92	2.68	3.49	3.16	2.56	3.50	3.08
June 1917	2.47	2.18	2.61	3.77	2.32	2.72	3.23	2.74
Average	2.71	2.39	2.89	3.39	2.60	2.76	3.18	2.84

(D) AVERAGE TOTAL RANGE IN CERTAIN TESTS COMBINED

	<i>Quartile</i>				<i>Tertile</i>			<i>Total</i>
	<i>I</i>	<i>II</i>	<i>III</i>	<i>IV</i>	<i>I</i>	<i>II</i>	<i>III</i>	
Eight Trabue	3.57	3.57	3.56	3.18	3.69	3.40	3.33	3.47
Six Mathematics	2.84	3.43	2.86	3.81	2.70	3.38	3.62	3.23
Five Directions	1.91	2.33	2.66	2.91	2.11	2.20	3.05	2.45
Three Reading	1.59	1.74	1.62	2.45	1.73	1.57	2.25	1.85
Average	2.48	2.77	2.68	3.09	2.56	2.64	3.06	2.75

(E) AVERAGE INTERQUARTILE RANGE (APPROXIMATION) IN THE ELEVEN TESTS

	<i>Quartile</i>				<i>Tertile</i>			<i>Total</i>
	<i>I</i>	<i>II</i>	<i>III</i>	<i>IV</i>	<i>I</i>	<i>II</i>	<i>III</i>	
February 1916	1.59	1.79	1.81	1.97	1.60	1.85	1.92	1.79
February 1917	1.42	1.71	1.73	2.34	1.50	1.63	2.27	1.80
June 1917	1.57	1.50	1.88	2.57	1.52	1.75	2.37	1.88
Average	1.53	1.67	1.81	2.29	1.54	1.74	2.19	1.82

Although the range is not so reliable as other measures of variability, still some deductions can be drawn from Table XIII which are significant. In all the parts of this table the fourth quartile shows consistently a marked increase in variability over the first, and likewise, the third tertile over the first, except in the case of the eight Trabue tests. This shows that among different abilities and in the same ability the range of achievements of the low ranking pupils is greater than that of those ranking high. Is this because of the poor showing, oftentimes almost absolute failure, they make in some tests? Parts B and C of Table XIII bear specifically upon this question. The pupils ranking low consistently have a greater range above their median achievement than do the pupils ranking high. Table XXIII shows that for all three testings there were twenty very low scores, almost absolute failures, in the highest tertile, and thirty-six very low scores in the lowest tertile. The difference between these two numbers is not sufficient to account for the greater range in ability on the part of the duller pupils. The results tend to show that the greater variability of the low ranking pupils is due to some factor inherent in the nature of their work.

Parts A, B, and C of Table XIII should be compared with

Individual Variability Compared with Group Variability 35

the first three columns of Table IV, which show the greatest total range of achievements in the second and third quartiles, and also a smaller range below the median in the fourth quartile than below the median in the first quartile. The data of Table XIII, as has already been pointed out, are more reliable than those of Table IV. The results shown in Table XIII will be discussed further in connection with the results of Tables XIV, XVIII, and XIX.

TABLE XIV

COMPARISON OF THE VARIABILITY OF INDIVIDUAL RANGES IN THE ELEVEN TESTS BY QUARTILES AND BY TERTILES

The per cents given are computed from the average of the average ranges in the three testings,—* Except the average of the Eight Trabue, Six Mathematics, Five Directions, and Three Reading Tests.

The table reads as follows: In the eleven tests the total range of Quartile II is 98 per cent of Quartile I, etc.

According to Values of Q	Per Cent Which the Variability of Each Quartile and Each Tertile is of Those Higher. (I is considered highest)								
	1	2	3	4	5	6	7	8	9
	Quartile						Tertile		
	II of I	III of I	IV of I	III of II	IV of II	IV of III	II of I	III of I	III of II
Total Range	.98	1.10	1.31	1.12	1.34	1.19	1.11	1.27	1.14
Range Above Median	1.15	1.17	1.42	1.02	1.23	1.21	1.19	1.34	1.13
Range Below Median	.88	1.07	1.25	1.21	1.42	1.17	1.06	1.22	1.15
*Average Range of Four Groups	1.12	1.08	1.25	.97	1.12	1.16	1.03	1.20	1.16
Inter-Quartile Range (Approximation)	1.09	1.18	1.50	1.08	1.37	1.27	1.13	1.42	1.26

The data of Table XIV are computed from the averages of Table XIII. This table summarizes the evidence on the question as to whether the duller pupils or the brighter pupils have the greater range in their achievements. Column 3 shows that on an average the lowest quartile had a range above the individual medians 42 per cent greater than the first quartile. Further, the range of the fourth quartile below the individual med-

ians was only 25 per cent greater than that of quartile one. However, these two figures, 42 and 25, should not be compared at face value. By the range of the tests used and the placement of median ability above the median of the range of the tests the possibility of large ranges above their medians was limited for the pupils ranking high, while in all other cases the range of the tests was sufficient to allow for the maximum individual range in either direction from the individual median. This would tend to make the 42 per cent increase in the range of the fourth quartile over the first somewhat greater than it should be.

The number of pupils making the highest score possible in the different tests shows that the range of ability covered was not an important factor in limiting the variability of the pupils ranking highest. In 21 of the 33 tests the highest score possible was reached by none of the pupils; in 9 it was made by a relatively small number; and in only 3 tests was the highest score possible made by a relatively large number of the pupils. That the range of ability covered by the tests was not an important factor in limiting the variability of the highest ranking pupils is shown further by certain results in Table XIV. The percentage of increase of the lower quartiles over the higher quartiles is as great in the case of the approximation of the interquartile range as in the total range. If the range of ability of the tests had been operative to any great extent it should have affected the total range of variability to a noticeably greater extent than the interquartile range.

Another point should be mentioned in this connection. In the range below the median there is undoubtedly a factor which is not present in the range above. Low scores in these tests are sometimes caused by external conditions,—chance occurrences such as the dropping of a pencil, becoming amused at some part of a rate test, etc., while high scores are not so caused. High scores are the result of ability; low scores are the result of either less ability or the failure of ability to function due to various causes. Thus in both cases the lower range is increased by this second factor which tends to make the percentage of increase in variability lower.

Allowing for these corrections the results seem to show that the low ranking pupils of this group are inherently more variable in

their achievement than the pupils ranking high. As to the reason for the one exception suggested above, namely, the eight Trabue tests combined in which the low ranking pupils are least variable, the investigation offers no evidence. It may be that the tests are better standardized, or that the ability required for these tests is more specific, or there may be some other reason for the results. It should be observed, however, that two problems are involved in this connection. One is the variability of individuals among different abilities and the other is the variability of individuals in different testings of the same ability. A large amount of variability among several traits or abilities does not necessarily imply great variability among several tests of the same trait. The Trabue scales test a single trait while the eleven tests cover several traits. Ability in each trait may remain about the same relatively from one testing to another and still there may be great variability among the several tests.

The amount of variability among the different tests cannot be compared directly with the amount among combined similar tests, shown in Section D of Table XIII, because in each case the number of tests combined is different from the number of different tests² in Section A of the table. This could be accomplished by some method of weighting but such will not be attempted here.

2. DISTRIBUTION OF INDIVIDUAL VARIABILITY

Several measures of the amount of individual variability have been found by taking different single measures of the range in achievement. The distribution of all the scores above and below the individual medians will throw more light upon this problem. Such distributions for the average of the three testings by tertiles and for the entire group by each testing are given in Table XV. The frequencies here are expressed in per cents so that the distributions for the eleven different tests may be compared later with the distributions for the combined similar tests. Figs. 16a to 17c show graphically the data in this table.

² These obviously are not all different tests in the sense of testing strictly different abilities. The two Trabue tests are of course for the same ability and the mathematics tests are for rather closely related abilities. The use of the phrase "eleven different tests" will be continued in the study with this limitation understood.

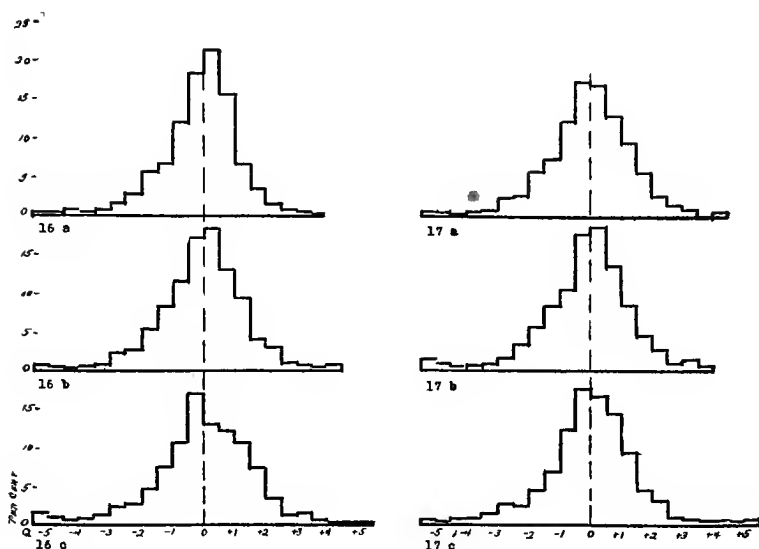
TABLE XV

DISTRIBUTION (IN PER CENTS) OF SCORES ABOVE AND BELOW THE INDIVIDUAL MEDIANS IN THE ELEVEN TESTS TRANSMUTED INTO MULTIPLES OF Q BY THE ORIGINAL DISTRIBUTIONS

The Measure of Central Tendency is the Median of the Individual's Scores Transmuted into Multiples of Q.

Value in Q	Average of the Three Testings (Feb. 1916, Feb. 1917, June 1917) by Tertiles:			Entire Number by Testings:		
	I	II	III	Feb. 1916	Feb. 1917	June 1917
+5.0 to —			.1			.1
+4.5 to +4.9			.1			.1
+4.0 to +4.4		.3	.1	.1		.3
+3.5 to +3.9	.1	.1	.4		.3	.4
+3.0 to +3.4	.4	.9	1.4	.9	1.1	.6
+2.5 to +2.9	.9	1.0	1.1	1.4	.8	.9
+2.0 to +2.4	1.5	3.0	3.7	2.5	2.7	3.0
+1.5 to +1.9	3.4	4.0	7.3	5.8	4.4	4.5
+1.0 to +1.4	6.7	9.5	10.6	9.2	8.4	9.1
+ .5 to + .9	15.7	13.0	12.1	13.0	13.5	14.3
.0 to + .4	21.3	18.2	13.0	17.0	18.8	16.7
.0 to — .4	18.4	17.2	17.0	17.2	17.9	17.5
— .5 to — .9	12.0	11.9	10.7	12.1	10.2	12.2
—1.0 to —1.4	6.9	8.2	7.2	7.3	8.1	6.9
—1.5 to —1.9	5.7	5.3	4.9	5.9	5.6	4.4
—2.0 to —2.4	2.8	2.5	2.9	2.8	3.2	2.3
—2.5 to —2.9	1.6	2.3	2.7	2.4	1.6	2.5
—3.0 to —3.4	.9	.8	1.1	.8	.6	1.4
—3.5 to —3.9	.5	.5	.8	.5	.5	.8
—4.0 to —4.4	.6	.4	.3	.1	.4	.8
—4.5 to —4.9	.3	.3	.9	.3	.8	.4
—5.0 to —	.3	.5	1.5	.6	1.1	.8

Figs. 16a, 16b, and 16c represent the distributions of the scores of Tertiles I, II, and III respectively. From the form of these curves it is evident that in the total distribution of their scores the pupils ranking lowest, those of the third tertile, are most variable in their achievements. The mode of the third tertile is not so pronounced as that of the first tertile. The range above the median of the third is greater than that of the first, and the range below the median of the third shows more extreme cases. All three curves tend to bring out the difference between the distribution of scores above the median and the distribution below. The range below is greater and more regular in its decline. The two halves of the curves show one similarity which is spurious. All the scores of 5Q or more are grouped into the last frequency because of the extreme range



Figs. 16a to 17c. Distribution (in Per Cents) of Scores Above and Below the Individual Medians in the Eleven Tests Transmuted into Multiples of Q by the Original Distributions.

Figs. 16a, b, c. Averages of the Three Testings by Tertiles

Fig. 16a. Tertile I

" 16b. " II

" 16c. " III

Figs. 17a, b, c. Entire Number by Testings

Fig. 17a. Feb. 1916

" 17b. Feb. 1917

" 17c. June 1917

in the graph which a few of the scores would have necessitated, —18.7Q in one case. This suggests a relation between the high and low scores of extreme variability which does not exist.

Figs. 17a, 17b, and 17c represent the distribution of scores above and below the individual medians in February, 1916, February, 1917, and June, 1917 respectively. Their significance is in their similarity. There is only one point of difference to note. It is the greater length of the curve in Fig. 17c when it nears the base line. The increase is not enough to be especially significant, and, moreover, it is in that part of the curve which is least reliable. However, it is in accord with the slight increase which was found in the Q of the third testing, and tends to show that these pupils became more variable in their own achievements the longer they remained in school.

TABLE XVI

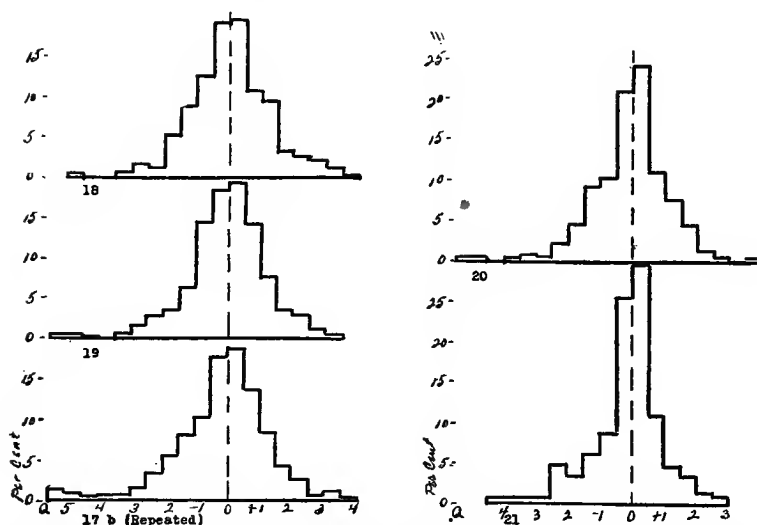
DISTRIBUTION (IN PER CENTS) OF SCORES ABOVE AND BELOW THE INDIVIDUAL MEDIAN IN CERTAIN TESTS TRANSMUTED INTO MULTIPLES OF Q BY THE ORIGINAL DISTRIBUTIONS

The Measure of Central Tendency is the Median of the Individual's Scores Transmuted into Multiples of Q.

Value in Q	Eight Trabue	Six Mathe- matics	Five Directions	Three Reading
+5.0 to —				
+4.5 to +4.9				
+4.0 to +4.4				
+3.5 to +3.9	.4		.3	
+3.0 to +3.4	1.1	.5		
+2.5 to +2.9	2.1	1.1	.6	.9
+2.0 to +2.4	2.6	3.0	1.4	1.4
+1.5 to +1.9	3.5	3.6	4.7	3.2
+1.0 to +1.4	9.7	7.9	7.8	4.2
+ .5 to + .9	10.9	14.1	11.1	10.6
.0 to + .4	19.7	19.7	24.1	29.6
.0 to — .4	19.2	18.5	21.0	25.5
— .5 to — .9	12.8	14.6	10.3	8.8
—1.0 to —1.4	9.0	6.5	9.4	6.0
—1.5 to —1.9	5.1	3.5	4.7	3.2
—2.0 to —2.4	1.2	3.0	2.2	4.6
—2.5 to —2.9	1.6	1.8	.6	.5
—3.0 to —3.4	.7	.9	.8	.5
—3.5 to —3.9			.3	.5
—4.0 to —4.4		.2		.5
—4.5 to —4.9	.4	.5	.3	
—5.0 to —		.5	.3	

Table XVI gives data for the four groups of combined tests which are similar to the data of Table XV for the eleven different tests. The frequencies of scores above and below the individual medians are expressed in percentages of the total number of scores in each combined group. Figs. 18 to 21 represent graphically the distributions of this table. Fig. 17b, representing the distribution for the eleven tests in February, 1917 is repeated in order to facilitate comparison.

These figures cannot be compared directly because, as has already been pointed out, the number of tests is different in each case. Expressing the frequencies in percentages equates the surfaces of distribution and permits some inferences to be drawn concerning the general shape of the curves. Figs. 18 and 19 representing the eight Trabue tests and the six mathematics tests are strikingly similar to Fig. 17b which represents the eleven tests. They do not show as much variability among the achievements of the individual as in the case of the eleven



Figs. 18 to 21. Distribution (in Per Cents) of Scores Above and Below the Individual Medians in Certain Tests Transmuted into Multiples of Q by the Original Distributions.

Fig. 18. Eight Trabue Tests Combined

Fig. 19. Six Mathematics Tests Combined

Fig. 17b. Feb. 1917 Testing (Repeated)

Fig. 20. Five Directions Tests Combined

Fig. 21. Three Reading Tests Combined

tests but they show a rather surprisingly large amount of variability. The mode is more pronounced in that the width of great density is larger. The extent of the curves and their shape near the base line are quite similar. The curves of Figs. 20 and 21 representing the five directions tests and the three reading tests differ from the others rather markedly. The smaller number of tests is probably a very potent reason for this, especially in the latter case.

TABLE XVII

THE Q OF THE DISTRIBUTION OF SCORES ABOVE AND BELOW THE INDIVIDUAL MEDIAN FOR THE THREE TESTINGS AND FOR CERTAIN TESTS COMBINED

Tests	Q
February, 1916	.81
February, 1917	.79
June, 1917	.80
Eight Trabue	.73
Six Mathematics	.71
Five Directions	.58
Three Reading	.46

The Q's of Table XVII were calculated from the distribution of scores above and below the individual medians shown in Tables XXXVII and XXXVIII in the Appendix. They should be compared with 1.00Q, the variability of the group used as the standard. The results here are slightly smaller than those of Table XI because the scores of the less variable pupils beyond the Q, but still less than the Q of the more variable pupils, reduce the size of the Q in the total distribution.

TABLE XVIII

DISTRIBUTION BY TERTILES OF RANGES ABOVE AND BELOW THE INDIVIDUAL MEDIANS IN THE ELEVEN TESTS IN VALUES OF Q

The Measure of Central Tendency is the Median of the Individual's Scores Transmuted into Multiples of Q.

Value in Q	Feb. 1916			Feb. 1917			June 1917			Average		
	Tertile			Tertile			Tertile			Tertile		
	I	II	III	I	II	III	I	II	III	I	II	III
+5.0 to +5.4									1			.3
+4.5 to +4.9									1			.3
+4.0 to +4.4		1					1	1			.7	.3
+3.5 to +3.9				1	1		1		2	.3	.3	1.0
+3.0 to +3.4	2	3	2	3	5		1	1	2	1.0	2.3	3.0
+2.5 to +2.9	2	2	4	2	1	1	3	2	2	2.3	1.7	2.3
+2.0 to +2.4	2	8	4	5	6	5	3	6	8	3.3	6.7	5.7
+1.5 to +1.9	9	8	10	4	5	6	5	4	5	6.0	5.7	7.0
+1.0 to +1.4	4	2	3	9	4	4	7	7	2	6.7	4.3	3.0
+ .5 to + .9	5		1	4	4	2	4	3		4.3	2.3	1.0
0 to + .4												
0 to - .4												
-.5 to -.9	1	1	2	1		2	2	2	1	1.3	1.0	1.7
-1.0 to -1.4	2	2	3	2	4	2	6	3	3	3.3	3.0	2.7
-1.5 to -1.9	5	7	6	8	6	2	2	3	6	5.0	5.3	4.7
-2.0 to -2.4	6	3	2	5	4	2	2	3	2	4.3	3.3	2.0
-2.5 to -2.9	5	5	4	2	5	4	4	3	4	3.7	4.3	4.0
-3.0 to -3.4	3		3		2	1	3	4	2	2.0	2.0	2.0
-3.5 to -3.9	1	2	1			4	3	2		1.3	1.3	1.7
-4.0 to -4.4	1			2	1		2	2	1	1.7	1.0	.3
-4.5 to -4.9		1	1	2	1	2			1	.7	.7	1.3
-5.0 to -5.4		1									.3	
-5.5 to -5.9						1		1	1		.3	.7
-6.0 to -6.4				1						.3		
-6.5 to -6.9		1				1		1			.7	.3
-7.0 to -7.4						3						1.0
-7.5 to -7.9												
-8.0 to -8.4									1			.3
-8.5 to -8.9									1			.3
-9.0 to -9.4			1		1						.3	.3
-9.5 to -9.9												
-10.0 to —	1	1	1	1				1		.3	.3	.7

Individual Variability Compared with Group Variability 43

TABLE XIX

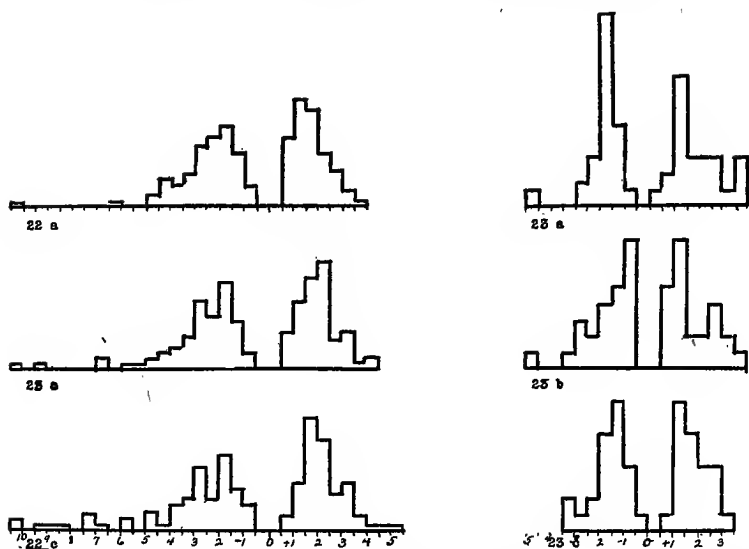
DISTRIBUTION BY TERTILES OF RANGES ABOVE AND BELOW THE INDIVIDUAL MEDIANS IN CERTAIN TESTS IN VALUES OF Q

The Measure of Central Tendency is the Median of the Individual's Scores Transmuted into Multiples of Q.

Value in Q	Eight Trabue			Six Math- ematics			Five Directions			Three Reading		
	Tertile			Tertile			Tertile			Tertile		
	I	II	III	I	II	III	I	II	III	I	II	III
+3.5 to +3.9	3	1							1			
+3.0 to +3.4	1	2	1		1	1						1
+2.5 to +2.9	3	4	4	2	2	1						1
+2.0 to +2.4	3	2	4	2	4	5		2	3			3
+1.5 to +1.9	3	2	6	5	2	4	4	2	8	4	2	4
+1.0 to +1.4	8	8	8	5	9	6	3	8	7	4	3	1
+ .5 to + .9	2	5	1	8	5	7	10	12	1	11	7	6
0 to + .4	1			2	1		7		2	8	12	8
0 to - .4			1	2	1	1	2	3	1	6	6	8
- .5 to - .9	1	8	4	6	4	5	4	9	8	6	6	5
-1.0 to -1.4	5	5	8	6	8	2	10	6	4	4	6	3
-1.5 to -1.9	12	4	6		4	6	3	5	2	4	1	2
-2.0 to -2.4	3	2	2	7	1	4	4		4	3	4	3
-2.5 to -2.9	2	3	1	3	2	2		1	1		1	1
-3.0 to -3.4		1	2		2	1	1		2	1		
-3.5 to -3.9					1							2
-4.0 to -4.4						1						
-4.5 to -4.9	1	1				1			1			
-5.0 to -5.4												
-5.5 to -5.9												
-6.0 to -6.4									1			
-6.5 to -6.9												
-7.0 to -7.4												
-7.5 to -7.9						1						
-8.0 to -8.4												
-8.5 to -8.9												
-9.0 to -9.4												
-9.5 to -9.9												
-10.0 to —					1							

Tables XVIII and XIX and Figs. 22a to 25c are introduced to supplement the data given in Tables XIII and XIV. "Nothing short of the entire distribution table is a complete measure of a variable fact. . . ."³ The first nine columns of Table XVIII are not separately represented graphically. The last three columns are the averages of the respective tertiles for the three testings. Figs. 22a, b, and c show these averages for the three tertiles of the group. The curves, of course, are bimodal

³ Thorndike, E. L., *Mental and Social Measurements*, p. 36.

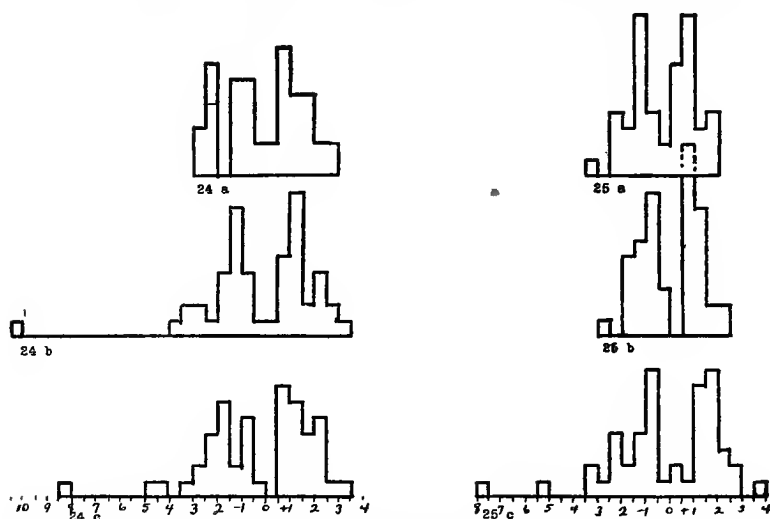


Figs. 22a to 23c. Distribution by Tertiles of Ranges Above and Below the Individual Medians in the Eleven Tests and in the Eight Trabue Tests in Values of Q. The Measure of Central Tendency is the Median of the Individual's Scores Transmuted into Multiples of Q.

Average of Three Testings
 Fig. 22a. Ranges in Tertile I
 " 22b. " " " II
 " 22c. " " " III

Eight Trabue Tests
 Fig. 23a. Ranges in Tertile I
 " 23b. " " " II
 " 23c. " " " III

because they represent two variables, the ranges above and below the median. They are joined to show the increase in the extent of the ranges of the pupils of the third tertile over those of tertiles two and one. The curves show the greater range of the extreme scores of the third tertile both above and below the median achievement, being especially true of the range below the median. Here again the curves are lopped off at $10Q$ and more minus. Finally, these curves show one point upon which Tables XIII and XIV do not give definite evidence. The 25 per cent increase below the median in the range of tertile three over tertile one is not accounted for chiefly by a very few extremely variant ranges but by the greater variability of this tertile in general. Further, the few extreme ranges above the median count still less in effecting the 45 per cent increase in the range of the third tertile above.



Figs. 24a to 25c. Distribution by Tertiles of Ranges Above and Below the Individual Medians in the Six Mathematics and Five Directions Tests.

Six Mathematics Tests
Fig. 24a. Ranges in Tertile I
" 24b. " " " II
" 24c. " " " III

Five Directions Tests
Fig. 25a. Ranges in Tertile I
" 25b. " " " II
" 25c. " " " III

Figs. 23a, b, and c represent similar data for the eight Trabue tests; Figs. 24a, b, c such data for the six mathematics tests; and Figs. 25a, b, c such data for the five directions tests. The same number of cases, twenty-four above and twenty-four below the median, is represented by the surface of each graph. The figures for the combined tests disclose fewer extremely variant ranges. Figs. 23a and 23c show that the exception to the increase in the range of the pupils ranking low over those ranking high, namely, in the eight Trabue tests, is not the result of a few extremely variant ranges in the first tertile, but an inherent result of the form of distribution.

The measures of extreme variability are emphasized not because they are thought to have ordinarily more significance than measures of variability near the central tendency but because it is one of the chief purposes of this investigation to study the extremely variant achievements.

The results of this topic and of the preceding one also tend to show that the pupils ranking lowest are most variable. Before leaving the topic further comparison of these results should

be made with the results obtained by ranks and given in Table IV. The data of Table IV show that the pupils ranking lowest are no more variable than are the pupils ranking highest, and that the pupils ranking nearer the median are the most variable ones. The average range of quartiles two and three is shown to be 16 per cent greater than the average range of quartiles one and four. The range of the fourth quartile below the median is shown to be less than half as great as the range of the first quartile below the median, while by the classification by the Q variability it is shown to be 25 per cent greater than the range of the first. Likewise with the rest of the results of this table.

Another point should be noted in this connection. Among this group of pupils there are no such types or pronounced extremes as are represented by Fig. 1, constructed from the classification by ranks. The piling up of scores at each end of the range as shown in Cases 60 and 92 is a spurious result of the method caused by the failure to retain the relative proportions of the original distributions.

This comparison is additional evidence that the method of evaluating achievements in terms of ranks from the highest to the lowest in the group does not produce as reliable results in connection with the different achievements as does the method used in this investigation.

3. OVERLAPPING OF DIVISIONS OF THE GROUP

There is another question in connection with this part of the problem of variability of the individual that should be asked concerning relative variability. Having single measures of the individual's variability and having the distribution of all his scores, the difference in ability of the different individuals should be known. Do the pupils who rank low and vary more in their achievements than the pupils who rank high, differ from those ranking high only a little in ability or do they differ a great deal? This difference can be measured by the per cent of overlapping of the scores among the different divisions of the group. Table XX gives the amount of overlapping of each quartile over the other according to three different points of reference,—the median, twenty-five percentile, and seventy-five percentile.

The three different testings, February 1916, February 1917, and June 1917, of Table XX are not sufficiently differentiated one from another to offer any points for special notice. Their overlappings are very similar. Consequently the average may be considered as typical of the three. The correspondence not the variation is the striking point of the overlappings revealed by the averages. The per cents of scores of Quartiles IV, III, and II that exceed the seventy-five percentile of Quartile I show a very close agreement with the per cents of Quartiles I, II, and III that extend below the twenty-five percentile of Quartile IV. The comparisons are: 4.7 with 3.5, 6.3 with 6.1, and 10.3 with 10.3. Similar comparisons using the per cents of the same quartiles above the median of the first and below the median of the fourth give: 9.1 with 9.3, 18.6 with 14.1, and 24.7 with 26.9. Other comparisons that might be made would disclose about the same agreement. This shows that the high scores of the low pupils overlap the high scores of the high pupils to an extent that corresponds very closely with the overlapping of the low scores of the high pupils over the low scores of the low pupils.

TABLE XXI

DIFFERENCE IN ACHIEVEMENT BETWEEN QUARTILES MEASURED IN TERMS OF THE Q VARIABILITY OF THE GROUP

The table reads: Between the medians of Quartiles I and II there were 22.2 per cent of the scores of Quartile I and 25.3 per cent of the scores of Quartile II, etc.

	<i>Per Cent of Scores Between the Medians of Quartiles:</i>		
	<i>I and II</i>	<i>II and III</i>	<i>III and IV</i>
Per cent of higher quartile overlapping lower..	22.2	18.2	23.1
Per cent of lower quartile overlapping higher..	25.3	15.5	19.9
Average overlapping	23.8	16.9	21.5
Value in Q	.95	.65	.84

The results given in Table XXI are calculated from the averages in Table XX. The following example illustrates the method. In Quartile I, 27.8 per cent of the scores are below the median of Quartile II. This leaves 22.2 per cent of the scores of Quartile I between its median and the median of Quartile II, etc. The values in Q are taken from a table of "P.E. Values

Individual Variability Compared with Group Variability 49

Corresponding to Given Per Cents of the Normal Surface of Frequency, Per Cents Being Taken from the Median.”⁴ The results show a slightly greater difference between the first and second and between the third and fourth quartiles than between the second and third. In median achievement the fourth quartile is 2.44Q below the first quartile. This shows that the variability of the fourth quartile is that of a distinctly lower grade of work.

There still remains another question of interest and importance, namely: On the basis of individual achievement in the eleven tests how far from zero ability in the traits measured are the different divisions of this group? The answer to this would round out this section of the problem. It would mean that any score of an individual could be related not only to his other scores and to the scores of other individuals of the group, but also that its absolute value could be determined. These absolute values could be determined for the scores of the tests that have been built by scaling achievements from the zero point, but for the others they could only be estimated. Therefore, this part of the problem will have to be left unanswered. This serves to emphasize the need for more tests scaled from zero for the problems in educational diagnosis.

For the purpose of individual diagnosis the variability of a test should be standardized either by grade or by age of the pupils. Having such a measure the scores of an individual could be compared by transmuting them into multiples of this variability without the labor involved in this investigation of determining a measure of variability by testing a group. A large number of cases would reduce the unreliability of the measure of variability to a very small amount and would make it possible to secure very reliable measures of the relative achievements of the individual.

Questions 3, 4, and 5 in the statement of the problem, concerning the amount of the individual's variability, the distribution of individual variability, and the variability of bright, mediocre, and dull pupils, have been considered in this section of the study. The average amount of variability of the sub-

⁴Trabue, M. R., *Completion-Test Language Scales*, Teachers College, Columbia University, Contributions to Education, No. 77, p. 38.

jects of this investigation for the three testings in the eleven tests used has been found to be eighty-two per cent of the variability of the group in the same tests. The variability in the last testing is slightly greater than in the first testing. The distribution of the achievements of the individual approximates the normal surface of frequency, the chief differences being a more pronounced mode and skewness downward from the median. On the basis of three equal divisions of the group the bright pupils are least variable and the dull pupils are most variable in their achievements. The distributions of the achievements for all three divisions have the same general form.

VI

EXTREME VARIABILITY IN INDIVIDUAL CASES

1. EXTREME VARIABILITY IN DIFFERENT TESTS

Mention has already been made of certain probable causes of low scores, such as distractions of the moment due to chance occurrences, and abnormal mental or physical condition of the individual at the time of the test. It has also been pointed out that these factors are not effective in producing high scores, or if effective at all, only to a very slight extent in comparison with their effect in causing low scores. The effect of chance happenings and abnormal conditions upon the achievement of the pupil can be ascertained to some extent by classifying the extremely variable or erratic scores and also the boys who make them, and by comparing the results from re-examination under more closely controlled conditions with the original achievements.

As can readily be seen from the chart in Fig. 15 there are no distinct types of scores or individuals. Therefore the line dividing extreme variability from the rest of the distribution must be arbitrarily drawn. A distance of $3Q$ from the individual medians was chosen for the location of this line. It was placed here because at about this point is the beginning of the second slow decrease in the normal curve of probability as characterized by a "slow-rapid-slow" decline in either direction from the median. It includes 47.8 per cent of the scores on either side of the median. Scores $3Q$ or more from the median in each direction will be called erratic either plus or minus. In Table XXII all the erratic scores in the three testings are classified by testing for each test and by total and average for each test.

The totals at the bottom of Table XXII show an increase in the number of erratic or extremely variable scores both plus and minus in each of the two later testings. Other things being equal this would show that the individual's abilities to achieve in these tests had increased at different rates. If the identical tests or tests of equated values had been used for the second and third

TABLE XXII

NUMBER OF SCORES IN THE DIFFERENT TESTS 3 Q OR MORE PLUS OR MINUS

Test	Scores 3 Q or More Plus			Scores 3 Q or More Minus			Number of Tests Given	Total No. of Scores 3 Q or More:			Average No. of Scores 3 Q or More.		
	Feb. 1916	Feb. 1917	June 1917	Feb. 1916	Feb. 1917	June 1917		Plus	Minus	Plus or Minus	Plus	Minus	Plus or Minus
Multiplication	1	3	*	4	2	.	2	4	6	10	2.0	3.0	5.0
Division		1	*	1	5	*	2	1	6	7	.5	3.0	3.5
Algebra Add. Subt.	*	*	6	*	*	2	1	6	2	8	6.0	2.0	8.0
Algebra Mult. Div.	*	*	1	*	*		1	1		1	1.0		1.0
Trabue (Both tests)	3	3	5	2		2	6	11	4	15	1.8	.7	2.5
Reading	1	1				3	3	2	3	5	.7	1.0	1.7
Visual Vocabulary	1			1		6	3	1	7	8	.3	2.3	2.7
Composition	1	3		1			3	4	1	5	1.3	.3	1.6
Spelling				6	4	7	3		17	17		5.6	5.6
Opposites				2	8	1	3		11	11		3.7	3.7
Mixed Relations				1	4	6	3		11	11		3.7	3.7
Easy Directions	1				4	5	3	1	9	10	.3	3.0	3.3
TOTAL	8	11	12	18	27	32		31	77	108			

* This type of test not given.

testings the absolute amount of gain in each could be determined. Such evidence would be more reliable than the evidence obtained, which measures the individual's increase in ability in relation to the rest of the group.

This increase in the number of erratic scores might also be accounted for by the piling up of scores at the mode to a greater extent in the later testings thus reducing the extent of the Q and thereby increasing the transmuted value of a deviation of the same absolute amount in all three testings. Table IX contains some evidence on this point, but not enough to decide it either way.

In the composition tests the Q in terms of the same scale is smaller in the last two testings than in the first. In the first testing there are two erratic scores, one plus and one minus; in the second testing there are three erratic scores; and in the third testing there are no erratic scores. In reading, the identical test, Alpha 2, Part II, was repeated in the second testing. The Q is slightly larger in the second testing than in the first testing and the number of erratic scores is the same. The reading test of the third testing was composed of different selections and therefore the Q can not be compared with the Q of Alpha 2. The number of words in the spelling tests was the same throughout. The first test was the easiest and has a smaller Q than the last, showing a greater piling up of scores, but still, this Q which is much less than that of the last test lacks one of producing as many erratic scores as there are in the last test. Opposed to these results the smaller Q of the second opposites test produces decidedly more erratic scores than the larger Q's of the other two tests. Other examples could be cited showing either result.

The results of Tables IX and XXII, in so far as they bear on this question, show that there was no marked reduction of variability caused by the repetition of the tests and therefore that the increase in the number of erratic or extremely variable scores in the later testings was not caused to a large extent by smaller Q's.

Table XXII shows that in all three testings there were 108 erratic scores. Of these 29 per cent were plus and 71 per cent were minus. This gives an average number of erratic scores

in each testing which is 4.5 per cent of the total number in each testing. That is, of every hundred scores four and one half were 3Q or more from the individual medians. It shows that the curves were skewed downward for in the normal surface of frequency only 2.2 per cent of items are beyond 3Q.

The last three columns of Table XXII give the average number of scores plus, minus, and plus and minus in each group of closely related tests. The first two of the three columns are the more significant. They show that in the rate tests practically all the erratic scores are minus. It would be expected that they would show more erratic scores minus than plus because distractions of the moment operate in this direction and affect rate tests most of all. Excepting the Algebra Addition and Subtraction test which was given but once and which, moreover, was in process of construction, spelling caused more erratic scores than any other test, and all of these were erratic in the minus direction.

The number of erratic scores resulting cannot be taken as a criterion for judging the unreliability of a test except in cases where scores are caused by chance happenings. Within limits the possibility for such results in a test would appear to have an inverse relation to the reliability of the test. The possibility of fine discrimination in achievements and the possibility for the functioning of a wide range of ability appear to be two factors which have a direct relation to the value of a test in educational diagnosis of the individual.

Table XXIII summarizes the results in the first half of Table XXII in a different way from that in which they are summarized in the last half of that table. It shows the erratic scores plus and minus by tertile and total for each testing. The significance of the table is in the increase in the number of erratic scores both plus and minus of the second tertile over the first and of the third tertile over the second. The numbers of erratic scores plus are 4, 10, and 17, and the numbers of erratic scores minus are 20, 21, and 36 for Tertiles I, II, and III, respectively. Of the total number of erratic scores, the per cent in each tertile is as follows: first tertile, 22 per cent; second, 29 per cent; and third, 49 per cent.

TABLE XXIII
NUMBER OF SCORES 3 Q OR MORE PLUS OR MINUS BY TERTILES AND
TOTAL FOR EACH TESTING

	Tertile I			Tertile II			Tertile III			Total
	Plus	Minus	Total	Plus	Minus	Total	Plus	Minus	Total	
Feb. 1916	2	5	7	4	6	10	2	7	9	26
Feb. 1917	0	7	7	4	5	9	7	15	22	38
June 1917	2	8	10	2	10	12	8	14	22	44
Total	4	20	24	10	21	31	17	36	53	108

2. EXTREME VARIABILITY OF DIFFERENT BOYS

The classification of erratic scores by test in which they occurred is only part of their description. Under this topic another part is given,—the classification by boys who made them. The following questions are considered: What per cent of the boys made erratic scores? Were there more or fewer boys who made erratic scores the longer they remained in school? If a boy has erratic scores in one testing what is the expectancy of his having erratic scores in one or both of the other testings? How do the high, median, and low ranking divisions compare as to the number of boys making erratic scores?

Table XXIV gives the number of boys making erratic scores in each test in one testing only and in all the combinations of testings. For example, the three boys counted in spelling in next to the last column of the table are not counted under the separate years. The results shown in this table are not different from what would be expected in the light of Table XXII. Spelling and the rate tests,—opposites, mixed relations, and easy directions—show the largest number of boys making erratic scores minus. In comparing these totals, division by the number of times the tests were given is implied. The only point that should be noted in connection with the number of boys

making erratic scores plus is the complete lack of such in the tests just mentioned,—spelling and the rate tests—except one case in easy directions.

TABLE XXIV

NUMBER OF BOYS HAVING SCORES 3 Q OR MORE PLUS OR MINUS IN EACH TYPE OF TEST IN EITHER ONE OR MORE TESTINGS

	3 Q or more plus							3 Q or more minus							Number of Tests Given		
	Feb. 1916 only	Feb. 1917 only	June 1917 only	F. '16, F. '17	F. '17, J. '17	F. '16, J. '17	F. '16, F. '17, J. '17	Total	Feb. 1916 only	Feb. 1917 only	June 1917 only	F. '16, F. '17	F. '17, J. '17	F. '16, J. '17		F. '16, F. '17, J. '17	Total
Multiplication	1	3	*					4	4	2	*					6	2
Division		1	*					1		4	*	1				5	2
Algebra Add. Subt.	*	*	6					6	*	*	2					2	1
Algebra Mult. Div.	*	*	1					1	*	*							1
Trabue (both tests)	3	3	5					11	2		2					4	6
Reading	1	1						2			3					3	3
Visual Vocabulary	1							1	1		6					7	3
Composition	1	3						4	1							1	3
Spelling								1		3	1		1	3		9	3
Opposites								1	7	1	1					10	3
Mixed Relations								1	3	5		1				10	3
Easy Directions	1							1		3	4		1			8	3

* This type of test not given.

TABLE XXV

NUMBER OF BOYS MAKING DIFFERENT NUMBERS OF SCORES 3 Q OR MORE PLUS OR MINUS IN ALL THREE OF THE TESTINGS. EACH BOY IS COUNTED ONLY ONCE

No. of Boys	No. of Scores
17	0
24	1
18	2
6	3
5	4
2	5

Table XXV shows the number of boys making different numbers of erratic scores in all three testings combined. The results show that 76 per cent of the boys made one or more erratic scores in all of the three testings. However, the impression given by this percentage is not quite fair. Too great a penalty is placed upon the making of one erratic score in any one of the three testings. This measure should be supplemented by the average of the three testings. Table XXVI shows that in the testings taken separately there were 25, 30, and 34 boys respectively who made erratic scores. These numbers give an average for the three testings of 41 per cent of the boys who made erratic scores.

Table XXVI also answers the question as to whether more or fewer boys made erratic scores the longer they remained in school, showing that in the second and third testings the number was increasingly greater. In February, 1916, 35 per cent made one or more erratic scores; in February, 1917, 42 per cent; and in June, 1917, 47 per cent.

Another question arises in this connection: Do the pupils who make erratic scores make more or fewer per pupil in the later testings? From the data of Tables XXIII and XXVI it is found that the number of erratic scores per pupil making erratic scores in the first testing is 1.04, in the second testing, 1.27, and in the third testing, 1.29. The rest of the data included in Table XXVI show the number of scores plus and the number minus made by every boy in each testing and in all three testings combined. The table reads: In February, 1916, 47 boys made neither plus nor minus erratic scores; 7 made one plus score each and no minus scores; 17 made one minus score each and no plus scores; and 1 made one plus score and one minus score.

Table XXVII analyzes the number of boys opposite each number of testings accordingly as they made only plus, only minus, or both plus and minus scores in the different testings. The last case at the bottom of the table is interesting. In one testing this boy made one or more erratic scores plus, but none minus; in another testing he made one or more minus, but none plus; and in the remaining one of the three testings he made both plus and minus erratic scores. The table shows that the

TABLE XXVI

NUMBER OF BOYS MAKING SCORES 3 Q OR MORE PLUS OR MINUS AND THE
NUMBER OF SCORES OF EITHER TYPE THAT EACH BOY MADE

Feb. 1916					Feb. 1917								
Minus	3 Q or More	Plus				Minus	3 Q or More	Plus					
		0	1	2	3			0	1	2	3		
	0	47	7				0	42	6	1			
	1	17	1				1	16	3				
	2						2	4					
	3					3							
June 1917					The Three Testings								
Minus	3 Q or More	Plus				Minus	3 Q or More	Plus					
		0	1	2	3			0	1	2	3	4	5
	0	38	7	1			0	17	7	1	1		
	1	19	2				1	17	9	1			
	2	3	1				2	8	3	2			
	3	1				3	1						
						4	3	1					
						5	1						

erratic scores made by one individual are not confined to one type. In the three testings 16 boys made erratic scores both plus and minus, and 7 of these made erratic scores both plus and minus in the same testing. There were 30 boys who made only minus erratic scores and 9 who made erratic scores in the plus direction only. Of the 72 pupils who were tested 76 per cent made one or more erratic scores; 42 per cent made erratic scores in one testing only; 22 per cent in two testings; and 12 per cent in all three testings. Of the boys who showed this amount of variability in their achievements in one testing, 46 per cent showed it again in either one or both of the other two

TABLE XXVII

NUMBER OF BOYS MAKING SCORES 3 Q OR MORE PLUS, MINUS, AND PLUS AND MINUS IN ONE OR MORE OF THE THREE TESTINGS. EACH BOY IS COUNTED ONLY ONCE

	<i>No. of Boys</i>	<i>Type of Scores Made</i>
None of the Three Testings	17	
One of the Three Testings	30 $\left\{ \begin{array}{l} 8 \\ 18 \\ 4 \end{array} \right.$	$\begin{array}{c} + \\ - \\ \pm \end{array}$
Two of the Three Testings	16 $\left\{ \begin{array}{l} 1 \\ 8 \\ 6 \\ 1 \end{array} \right.$	$\begin{array}{cc} + & + \\ - & - \\ + & - \\ + & \pm \end{array}$
All Three Testings	9 $\left\{ \begin{array}{l} 4 \\ 2 \\ 1 \\ 1 \\ 1 \end{array} \right.$	$\begin{array}{ccc} - & - & - \\ - & - & + \\ - & - & \pm \\ + & + & - \\ + & - & \pm \end{array}$

testings. These figures show that a large percentage of the boys made extremely variable scores,—scores of 3Q or more above and below their median achievements.

One more question asked at the beginning of this topic remains to be answered: How do the high, median, and low divisions compare in the number of boys making erratic scores? The answer could be predicted from Table XXIII. Table XXVIII gives the facts. In every testing the second tertile has more boys making erratic scores than does the first tertile, and in every testing the third tertile has most of all, with one exception, February, 1916, when there were more in the second tertile. The totals show a consistent increase in the number. Using the average number of boys who made erratic scores it is found that 24 per cent are in the first tertile, 32 per cent are in the second tertile, and 44 per cent are in the third tertile.

TABLE XXVIII

NUMBER OF BOYS HAVING SCORES 3 Q OR MORE EITHER PLUS, OR MINUS;
OR PLUS AND MINUS BY TERILES AND TOTAL FOR EACH TESTING

	<i>Tertile I</i>				<i>Tertile II</i>				<i>Tertile III</i>				<i>Total</i>			
	<i>Plus only</i>	<i>Minus only</i>	<i>Plus and Minus</i>	<i>Total</i>	<i>Plus only</i>	<i>Minus only</i>	<i>Plus and Minus</i>	<i>Total</i>	<i>Plus only</i>	<i>Minus only</i>	<i>Plus and Minus</i>	<i>Total</i>	<i>Plus only</i>	<i>Minus only</i>	<i>Plus and Minus</i>	<i>Total</i>
Feb. 1916	1	4	1	6	4	6		10	2	7		9	7	17	1	25
Feb. 1917			6	6	3	4	1	8	4	10	2	16	7	20	3	30
June 1917	1	7	1	9	1	9	1	11	6	7	1	14	8	23	3	34
Total *	2	17	2	21	8	19	2	29	12	24	3	39	22	60	7	89

* In the totals for the three testings the same boy may be counted more than once.

3. REDUCTION OF VARIABILITY BY RE-EXAMINATION

In the preceding topic of this section it has been found that of all the erratic scores made 71 per cent were minus and 29 per cent were plus, and that of all the boys making erratic scores 16 per cent made plus scores, 55 per cent minus scores, and 29 per cent made both plus and minus scores. The problem here is to determine the reduction in the number of erratic scores in these same tests which a special examination under closely controlled conditions would produce.

The tests used in the special examination were identical with those used in the original testings. They were given about three weeks after the third testing. The time allowed was as nearly equal to the time in the original testing as was possible. Especial care was taken to insure the subject's best reaction in accord-

TABLE XXIX

COMPARISON OF SCORES IN ORIGINAL AND SPECIAL TESTS OF CERTAIN BOYS
HAVING SCORES 3 Q OR MORE MINUS IN ORIGINAL TESTS

<i>Tests</i>	<i>No. of Boys in Special Tests</i>		<i>No. of Scores in Tests</i>	<i>Scores more than 3 Q Minus</i>	<i>Scores less than 3 Q Minus</i>	<i>Per Cent of Scores more than 3 Q Minus</i>	<i>Reduction in Per Cent of Scores more than 3 Q Minus</i>
Spelling	3	Original	9	8	1	88.9	33.3
		Special	9	5	4	55.6	
Opposites	4	Original	12	5	7	41.7	4.2
		Special	8	3	5	37.5	
Mixed Relations	5	Original	15	5	10	33.3	33.3
		Special	10	0	10	0	
Easy Directions	4	Original	12	5	7	41.7	29.2
		Special	8	1	7	12.5	

ance with the directions of the test. These tests are described in Section III under Special Testing. Re-examination of all the boys who made erratic scores in the tests in which they made them would have produced the most reliable results. This, however, was inexpedient and consequently only a part of the group were re-examined.

In Table XXIX the results secured in the four tests used in the special examination are compared with the results of the original testings. The values in Q for all the tables in this topic were calculated by using the Q of the original distributions. Since the number of scores obtained in the special testing is not the same in every case as in the original testings and also because the number is not the same in all tests, the gains made are expressed in per cents. These are shown in the last column

of the table under Reduction in Per Cent of Erratic Scores. The special examination produced a reduction of 22 per cent in the number of erratic scores on the basis of the total number of scores made. All of this reduction of variability should not be credited to the elimination of accidental or unusual occurrences. Some of it is probably due to improvement through practice, especially in the case of the tests which had been used in the third testing.

TABLE XXX

COMPARISON OF SCORES OF CERTAIN BOYS IN SPECIAL TESTS WITH THEIR SCORES IN CORRESPONDING ORIGINAL TESTS

Test	No. of Boys in Special Tests	No. of Scores in Special Tests	Difference Between Score in Special Test and Original Test			Difference in Value in Q Between Special and Original Test			Composite Rank in Ability by Three Testings	Composite Rank in Variability by Three Testings
			Gain		Loss	Gain		Loss		
Spelling	3	9	7 18 16	0 8 3	5 3 5	4.3 1.3 10.0	0 2.1	1.0 1.0	1.5 58 65	59 72 56.5
Opposites	4	8		4 1 4 0	1 1 1 3		2.6 .3 2.6 0	2.1 1.3 2.1 1.9	60 15.5 69 48.5	44 46 50 65
Mixed Relations	5	10	13 7 11 2	1 8 13 9	1 1 1 4		4.4 2.3 3.8 .7	.4 2.7 .4 4.4 3.0	35 71 15.5 67 14	61.5 45 56.5 55 49
Easy Directions	4	8	6 4 4 1	7 8 8 5			2.5 1.7 1.6 .4	4.7 5.4 5.3 3.4	72 64 41 68	69 26.5 43 71

Table reads: In spelling 3 boys were re-examined in three tests each. One boy made a score 7 points higher than his score in February, 1916, a score equal to his score in February, 1917, and a score 5 points lower than his score in June, 1917. In values of Q his scores were 4.3 Q higher, equal, and 1.5 Q lower than his respective original scores. This boy ranked 59 in ability and 63.5 in variability. (1 being least variable.)

Table XXX analyzes by individuals the difference in value of Q between the original and special testings. It shows that the number of gains in score greatly exceeds the number of losses. There are 26 cases of gain and 7 of loss and 2 cases where the score is the same as in the original testing. From the results in this table it is found that the average gain in points over the original scores is 5 in spelling, 1.1 in opposites, 5.9 in mixed relations, and 5.6 in easy directions. These results are supplementary to the results of Table XXIX. Of the sixteen individuals re-examined all but one were below the median in ability and all but one were more variable than the median.

4. THE CAUSES OF EXTREME VARIABILITY

The results of this investigation show a rather large amount of variability among the different achievements of the individual when compared with the variability of the group. By the method described in Section IV the amount of this variation has been measured and the results have been given in Section V. In this section the extreme cases of variability,—those 3Q or more from the median scores of the individual—have been segregated and classified by tests and by individuals. Also the results from a re-examination of certain boys making extremely variable scores have been compared with the results from the original tests of the same boys. Some evidence concerning the causes of extremely variable or erratic scores has appeared in connection with other parts of the problem. Under this topic such evidence will be collected and some additional data will be discussed.

The following causes appear to be factors which may operate individually or in combinations to produce extremely variable or erratic scores:

- a. The nature of the tests used.
- b. The administration of the tests.
- c. Accidental or unusual occurrences.
- d. Statistical treatment of results.
- e. The ability of the individual in different traits.

The effect of these causes individually can be discussed from a general standpoint, but the extent to which each one operated

in producing specific erratic scores can not be definitely determined from the data of this investigation.

The tests used are well standardized in degrees of difficulty. However, the amount of the increasing increments of difficulty which can be shown by the individual score and also the range of ability which is covered vary considerably. In group measurements increments smaller than the interval of the scale can be measured by interpolation within the interval of the scale, but in measurements of the individual such calculations cannot be made and consequently increments smaller than the interval of the scale cannot be measured.

In transmuting the original scores into multiples of Q the value of a score was taken as one half interval higher than the actual score in all tests except composition in which case the value taken was the exact score. The values taken thus represent the midpoint of all the unmeasured achievements. Consequently the maximum displacement that could be produced by the interval of the scale is just barely less than one half the amount of Q representing the interval of the original distribution. The amount of Q representing one half the interval of each of the original distributions is shown in Table XXXI.

TABLE XXXI
THE AMOUNT OF Q REPRESENTING ONE-HALF THE INTERVAL OF THE DISTRIBUTIONS OF THE DIFFERENT TESTS

	<i>Feb.</i> 1916	<i>Feb.</i> 1917	<i>June</i> 1917
Woody Multiplication	.24	.42	
Woody Division	.19	.75	
Hotz Algebra, Add. and Subt.			.25
Hotz Algebra, Mult. and Div.			.19
Trabue B, J, L.	.30	.41	.46
Trabue C, K, M.	.38	.35	.25
Reading Tests	.12	.11	.16
Visual Vocabulary	.14	.04	.12
Composition	.09	.12	.10
Spelling	.32	.13	.16
Opposites	.27	1.07	.33
Mixed Relations	.11	.17	.17
Easy Directions	.21	.34	.66

Table XXXI shows the extent to which the different tests may have failed to record the achievements of the individual due to the extent of the interval of the distribution when transmuted into Q . Obviously an allowance for this can not be made for any given score because the known value nearest the

achievement has already been taken. The table shows that in only a few of the thirty-three tests could the interval of the distribution be of much significance in causing extremely variable scores. It may effect not only the extremely variable scores but also all other scores to the same extent. The point is that if it happens to act upon an achievement that should be just less than three Q in either direction from the median of the individual's achievement it puts that score in the group called erratic in this study. The effect of the interval of distribution is significant only in connection with particular scores; it does not affect theoretically either the average amount of variability found in Section V or the total number of erratic scores found in Section VI.

Another phase of the first factor in the causation of erratic scores is the range of ability covered by the test. The original distributions show that the upper limit of the scale was not reached by any of the pupils in the following tests: arithmetic, algebra, language, reading, visual vocabulary, and composition. In the spelling tests the upper limit was reached by seven, one, and eight pupils in the three testings respectively. In the association tests only a few pupils reached the upper limit except in the opposites test of February, 1917 and the easy directions tests of February and June, 1917 when a relatively large number of the pupils made the highest score possible. These tests did not permit the best pupils to show their ability in comparison with the rest of the group. Consequently the nature of the tests is a factor that probably prevented some extremely variable scores in the plus direction from the median.

The distributions of scores resulting from ranges insufficient to cover the ability of all the pupils increased to some extent the amount of Q representing the value of the score. This has been illustrated by Fig. 14 and discussed under Topic 2 of Section IV. It was pointed out there that the skewness of the curve tends to reduce the extent of the Q of the original distribution from what it would be in a normal distribution. This increases the variability of the scores when expressed in multiples of the Q which is smaller than it would be in a normal distribution. The effect is cumulative so that the farther a score is from the median of the original distribution the greater is

the amount of spurious variability produced by this cause. On account of its cumulative effect this cause probably placed some scores in the group called erratic and did not have the opposite effect on other scores. From the distributions it would appear that skewness could be effective to a marked degree only in the minus direction and only in the opposites test of February, 1917 and the Easy Directions tests of February and June, 1917. The same cause could operate to make a score less than 3Q from the individual median when it should be more than 3Q from it if the median achievement of the individual is more than 3Q from the median of the group. However, there are no such cases in this investigation. The form of the distribution as a cause of extreme variability on the part of the individual may be attributed partly to the nature of the tests and partly to the statistical treatment of the results.

The extent of the interval, the range of ability covered by the tests, and the form of distribution of the scores have been considered in connection with the nature of the tests as possible causes of extreme variability. The extent of the interval of the distribution was found to have but little effect in causing erratic scores, and since it has a compensating effect its significance is almost negligible. The range of ability covered by the test probably prevented some scores from being erratic in the plus direction. The form of distribution of the scores in certain tests tends to magnify the amount of variability in the minus direction, and probably makes the number of scores in the minus direction larger than it should be. The last two causes probably account to some extent for the disparity between the number of extremely variable scores in the plus direction and the number in the minus direction.

The practical significance of these causes of variability is illustrated by the use of standardized tests in the classification of pupils for the purposes of instruction. Here the individual rather than the group is the unit to be dealt with. In order to classify the individuals of a group with adequate exactness the range of the tests should be somewhat greater than the range of ability, and the interval between scores should be small.

The extent to which the administration of the tests caused erratic scores will have to be judged by the conditions under

which the tests were given. In the first testing five of the eleven tests were given at the schools from which the pupils came. The remaining six tests were given at the Speyer School. The tests were administered by several graduate students under the direction of Professor Briggs. The second testing was conducted by Dr. Fretwell, who was one of the group giving the tests the first time, and by the writer. The third testing was conducted by the writer. Five tests of the first testing were given in the regular class rooms of the public schools. The remaining six tests of the first testing and all the tests of the last two testings were given in the regular class rooms of the Speyer School. The tests were given during scheduled periods of the school day. Pupils were tested in regular class groups of about twenty-five each. Instructions concerning the tests were brief and of similar nature for each test at each succeeding testing. Considering these conditions it is probable that the administration of the tests had but little effect in producing erratic scores.

Accidental or unusual occurrences probably had a marked effect upon the scores of certain individuals. In one of the easy directions tests, for example, the completion of a face by the addition of the nose caused unusual merriment for certain pupils who by chance or design produced a rather grotesque face by the type of nose added. In a few cases this diversion impeded materially the speed of the work thus precluding a normal achievement.

In one or two instances in the spelling tests pupils, either because of some accidental occurrence or because of slowness, fell behind the rate at which the words were being pronounced. This probably reduced the number of words spelled correctly. It is probable, however, that if any extremely variable scores were produced by accidental occurrences practically all were in spelling and in the rate tests.

One bit of evidence bearing upon this cause and also upon the administration of the tests is found in connection with Table XXIX concerning the amount of reduction in per cent of erratic scores by special testing. When tested in groups of from two to five in spelling and in the rate tests, certain pupils who had made extremely variable scores in the original testings reduced their number of erratic scores by from four to thirty-three per cent of the total number of scores. Since the special tests

were given about three weeks after the third testing it is probable that practice entered into the reduction of these per cents to some extent. Apparently accidental occurrences had little effect upon the difficulty tests.

The statistical method of combining the results was chosen because it seemed to "preserve all the refinement of the original measurements" to a greater extent than other methods. That the variability of certain scores has been magnified to some degree by this method has already been pointed out in connection with the nature of the tests. However, the method used is certainly only a small factor in the causation of extremely variable scores.

Consideration of the ability of the individual as a factor in the causation of extremely variable scores involves a study of the individual's variability from an angle slightly different from the attack made thus far. The problem of chief concern has been the variability of the individual from his own median achievement. The problem presented now is the variability of the individual in his own achievements in the same or similar tests at succeeding testings. If exactly the same tests had been used at each succeeding testing this variability could be measured in terms of absolute amounts of gain and loss in the different tests. Since the identical tests were not repeated in all cases such variability must be measured in a different way. This can be accomplished either by finding the difference between the Q values of the scores in similar tests for the different testings, or by ranking the individual's achievements in each testing and finding the variation in ranks of the scores in question for the three testings.

If the scores which vary greatly from the individual's median achievement are mere chance happenings among the total number of achievements there should be a greater variation among the ranks in the three testings of the abilities in which such scores occur than among the ranks of the abilities which have no scores at so great a distance from the median. This of course would not hold if the ranking of all the individual's achievements is caused by mere chance, but even if this is true the rank among the individual's achievements of the abilities having scores at the greatest distance from the median would have the same cause as the rank of any other ability and should not

vary in rank any more than the abilities having no scores at the extreme distance from the median.

The variability of the individual in his own achievements in the three testings has been found for nine of the pupils most variable in the range of their scores in the first two testings and for eight of the pupils who had but one score of 3Q or more in the first two testings, and also for eight of the pupils most variable in all three testings, and eight who had but one score of 3Q or more in all three testings. The results are given in Table XXXII. In the treatment of the first two testings the two Trabue tests are combined because of their similarity thus making ten scores to be ranked. In the treatment of the three testings the two Trabue tests are combined, and the arithmetic and algebra tests are omitted because they are not similar enough to be comparable, thus making eight scores to be ranked.

The results given in Table XXXII show that the variation of the scores at the greatest distance from the individual's median achievement is not essentially different from the variation of those nearer the median. They vary in rank about the same as the other achievements. The last two columns of the table show further that the variation of the erratic scores in the rankings of the pupils having only one erratic score is no greater than the variation in the rankings of the pupils who have the most erratic scores. Inspection of charts like the one shown in Figure 15 shows that the absolute amount of variation of these extreme scores is greater than the variation of those nearer the median, but, like the items near the extremes of any distribution, they normally would be expected to vary more in absolute amount. The interpretation of these results is that in the ranking of the individual's achievements the variability of the extreme scores is caused by mere chance to no greater extent than the variability of the scores nearer the median.

From one point of view this topic,—the causes of extreme variability, is a study of the forces that prevent perfect correlation between different testings of the same ability, and from the same viewpoint the whole investigation is a study of the lack of perfect correlation between abilities and between different testings of the same ability. If there were perfect correlation in the above cases there would be, of course, no variability among either the achievements of the individual in different tests or

TABLE XXXII

THE VARIABILITY OF CERTAIN INDIVIDUALS IN THE RANKING OF THEIR OWN ACHIEVEMENTS FOR THE DIFFERENT TESTINGS

	9 of the Most Variable Pupils in the First Two Testings	8 Pupils Having One Erratic Score Each in First Two Testings	8 of the Most Variable Pupils in All Three Testings	8 Pupils Having One Erratic Score Each in All Three Testings
Number of differences found between ranks one or more of which are erratic.	16	8	39	24
Number of differences found between ranks none of which is erratic.	74	72	153	168
Average difference between ranks one or more of which are erratic.	2.6	2.9	1.9	1.6
Average difference between ranks none of which is erratic.	2.4	3.3	2.0	2.3
(Average difference between such number of ranks by chance ranking.)	3.3	3.3	2.6	2.6
Median difference between ranks one or more of which are erratic.	1.5	2.3	1.4	1.3
Median difference between ranks none of which is erratic.	2.1	3.3	2.1	2.3
(Median difference between such number of ranks by chance ranking.)	2.4	2.4	1.8	1.8

Explanation of the table. The scores of 9 of the most variable pupils were ranked from 1 to 10 in each of the first two testings. The differences between the ranks in similar tests were found,—90 in all. Of these 16 are differences between ranks of scores one or both of which are erratic and 74 are between scores neither of which is erratic. The average difference between the former scores is 2.6 ranks and between the latter is 2.4 ranks. The median difference between the scores one or both of which are erratic is 1.5 ranks and between the scores neither of which is erratic is 2.1 ranks.

among the ranks in the same or similar tests at different times. Such relation, however, has not been found to exist. Since variability is found to exist the question as to how constant it is logically follows. Forty-five coefficients of correlation between rankings of individual achievements in different testings were found. They range from $+.87$ to $-.29$ and average $+.32$ but are not considered to have enough significance to be introduced as evidence of any relation.

TABLE XXXIII
TEACHERS' RATINGS ON CERTAIN POINTS CONCERNING THE WORK
OF FIFTEEN PUPILS

Ind.	1 <i>Preparation of Lessons</i>			2 <i>Class Dis- cussion</i>		3 <i>Atten- tion in Class</i>		4 <i>Con- duct in Class</i>		5 <i>Inter- est in Sub- ject</i>		6 <i>Total</i>			7 <i>Vari- ability in Tests by:</i>		8	9	
	Consistent	Normal	Erratic	Consistent	Normal	Erratic	Consistent	Normal	Erratic	Consistent	Normal	Erratic	Consistent	Normal	Erratic	Range	Approximation of interquartile range	No. of scores erratic	No. of teachers suggesting reasons for variability
Least variable																			
42	2	2		2	1	1	2	2		3	1		12	7	1	1	20		2
17	2	1	1	2	2	1	3	1		3	1		12	5	2	3	1		2
64			4	2	1	1	1	2	1	1	1	2	5	6	9	5	14		3
59	1	3		1	3		1	3		3	4		3	17		6	3		
26	3	1		3	1		3	1		3		1	15	4	1	7	16		1
Tot.	8	7	5	10	8	2	10	9	1	10	9	1	48	39	13				
Near median																			
46	3	1		2	2		4			4			16	4		33	33.5	2	1
58	3		1	3	1		4			4			18	1	1	35	29.5		2
73	3	1		3	1		3	1		4			17	3		36	47	1	1
5	2	2		2			2	2		1	1	2	3	7	10	39	43	3	2
45	4			1	3		4			1	3		2	18		40	41	2	4
Tot.	6	6	8	7	7	6	8	5	7	8	6	6	37	31	32				
Most variable																			
61	1	3		3	1		2	2		2	2		10	10		65	46	2	2
29	2	1	1	2	2		1	2	1	1	3		8	9	3	68	60	2	1
19	2	1	1	1	2	1	1	2	1	1	2	1	6	9	5	69	72	5	1
52	1	2	1	1	3		2	2		1	3		7	11	2	70	68	4	1
66	2	1	1	2	1	1	2	2		3	1		12	6	2	71	71	5	2
Tot.	7	6	7	6	11	3	6	10	4	6	11	3	33	45	22				

Further information concerning the permanency of individual variability was sought by asking the teachers of these boys to rate some of them on certain points. Table XXXIII gives the results of the ratings by four teachers who had had all of the group selected in their classes for a considerable length of time. The following instructions were given to the teachers: "Check in the proper column your judgment of the following boys as to whether they have been consistent, normal, or quite erratic in the traits noted on the accompanying form. Add any remarks that will help to explain the character of their work." Normal was defined to mean the amount of variation that is normally expected. In the form given to the teachers the pupils were listed in groups of three, each group containing one of the least variable according to the tests, one of the most variable, and one near the median in variability. The findings concerning variability in the tests were unknown to the teachers. In Table XXXIII the least variable, the most variable, and those near the median in variability according to the tests are segregated. The table reads as follows: Individual No. 42 was judged consistent twice and normal twice in preparation of lessons, etc. Columns 6 to 9 inclusive were not in the form given to the teachers. In column 6 the teachers' judgments are summarized; column 7 shows the variability of the pupil as indicated by the range and approximation of the interquartile range in the tests; column 8 shows the number of the individual's scores 3 Q or more from his median in all testings; and column 9 shows the number of teachers suggesting reasons for the boy's variability.

The totals in column 6 of the table show that there is a slight tendency for the ratings of the teachers to agree with the rankings by the tests in the matter of variability. It is so slight however that it has but little significance. The last column of the table shows that the teachers suggested reasons for extreme variability in all three groups offering about as many for one group as for another. The following causes for erratic work were mentioned most frequently: physical defects, nervousness, absence, home life, and outside work. In order to get more definite information concerning the permanency of individual variability a systematic checking in certain points would have to be continued by the teachers over a fairly long period of time.

It should be pointed out in connection with this section of the study that the individuals making these extreme scores, both those making them above and those making them below their medians, offer much opportunity for further investigation by repetition of these same tests and other tests, and also by tests of their physical as well as their mental attainments. The time and labor necessary for such a study^o precluded the possibility of incorporating it in this investigation.

Answers to the following questions asked in connection with the problem have been proposed in this section of the study.

Question 6. To what extent are there extremely variable or erratic scores? The line marking off erratic scores was arbitrarily placed at a distance of 3Q in each direction from the individual medians. On this basis the average number of erratic scores in all three testings was found to be 4.5 per cent of the total number of scores for each testing. The later testings show an increase in the number of erratic scores. The number in the second testing is 46 per cent larger than the number in the first testing, and the number in the third testing is 16 per cent larger than the number in the second testing.

Question 7. How do the bright, the mediocre, and the dull pupils compare as to the number who make erratic scores, and as to the number of such scores each one makes? The extreme scores of this group of pupils are not especially characteristic of any one division of the group. Forty-four per cent of all the boys who made erratic scores are in the third tertile; 32 per cent are in the second tertile; and 24 per cent are in the first tertile. Using the average number of pupils making erratic scores and the average number of erratic scores made, it is found that the average number of erratic scores per pupil is 1.14 in tertile I, 1.07 in tertile II, and 1.36 in tertile III.

Question 8. What are the causes of the extremely variant scores? The causes of the scores varying 3Q or more from the individual's median achievement, in so far as this study throws light upon them, have been analyzed under the five headings: the nature of the tests used; the administration of the tests; accidental or unusual occurrences; statistical treatment of the results; and the ability of the individual in different traits.

The nature of the tests used probably prevented some cases

of extremely variable scores in the plus direction from the individual's median which would have appeared if the range of ability covered by the tests had been greater. The nature of the tests and the statistical treatment of the results seem to have magnified the amount of variability of a relatively small proportion of the scores. The administration of the tests, in so far as it can be judged by the conditions of the testings, had practically no effect upon the variability of the scores. Accidental or unusual occurrences probably caused a few erratic scores. Under a more detailed administration of the tests such occurrences and their effect could be definitely accounted for in the results. From the evidence of this study it appears that the ability of the individual is the greatest of the five factors in the causation of scores which vary 3Q or more from his median achievement.

VII

CORRELATION BETWEEN MEASURES OF ABILITY, MEASURES OF VARIABILITY, AND MEASURES OF ABILITY AND VARIABILITY

1. CORRELATION BETWEEN MEASURES OF ABILITY

The results that have been set forth up to this point have dealt with variation. They may be considered as showing certain positive relations, but in an indirect way. In this section of the investigation different relations will be studied by means of coefficients of correlation. The last question in the statement of the problem will be considered. This question concerns the relation between different measures of ability, the relation between different measures of variability, and the relation between measures of ability and variability.

The first coefficients that are given are between different methods of ranking pupils for composite achievement. Three methods were used. First, each one of the seventy-two pupils was ranked by the average of his eleven ranks. That is, the pupils were ranked from one to seventy-two in each test. The eleven ranks of each pupil were then averaged and these averages were ranked from one to seventy-two, the smallest being ranked one. The second method was the same as the first except that the median rank was used instead of the average rank. The third method was by median rank in the eleven tests as obtained from the scores transmuted into multiples of Q . Using the median of the individual's ranks in values of Q the pupils were ranked from one to seventy-two as in the other methods.

Three correlations were then calculated between the rankings by each method,—February 1916 with February 1917; February 1917 with June 1917; and February 1916 with June 1917. The coefficients of these correlations are given in Table XXXIV. In calculating all the coefficients in this section the formula

$$\rho = 1 - \frac{6 D^2}{n(n^2-1)}$$

was used. The value of these coefficients in terms of the Pearson r has been inferred from a table¹ of such values. In all cases the inferred value of the coefficient is given. The unreliability of the coefficients was determined by the formula

$$\text{P.E.} \quad t.r - \text{obt. } r = .6745 \frac{1 - r^2}{n}$$

"The probable divergence of the true coefficient of correlation from that obtained from a limited random selection of related pairs, is a variable fact with a mode at 0, and a variability which serves as the measure of the unreliability."² The P.E. is the measure limiting the fifty per cent of this variability which is nearest the coefficient obtained.

TABLE XXXIV
CORRELATION BETWEEN COMPOSITE RANKINGS IN ABILITY

		P.E. of r
Average Rank by Rank in Eleven Tests		
	Feb. 1916 with Feb. 1917.....	.77..... .03
	Feb. 1917 with June 1917.....	.78..... .03
	Feb. 1916 with June 1917.....	.68..... .04
Median Rank by Rank in Eleven Tests		
	Feb. 1916 with Feb. 1917.....	.69..... .04
	Feb. 1917 with June 1917.....	.73..... .04
	Feb. 1916 with June 1917.....	.53..... .06
Median Rank by Values of Q in Eleven Tests		
	Feb. 1916 with Feb. 1917.....	.69..... .04
	Feb. 1917 with June 1917.....	.69..... .04
	Feb. 1916 with June 1917.....	.54..... .06

The method of ranking the pupils by their average achievement gives distinctly higher coefficients of correlation than either of the other methods. The results obtained by ranking them in ability by the median of their eleven ranks agree very closely with the results obtained by ranking them by their median rank in values of Q. Coefficients of correlation obtained from both of these methods are approximately 10 per cent lower than the coefficients obtained from the method by average rank. The reason for the difference between the coefficients obtained from the rankings by the average of the eleven ranks and the coefficients obtained from the rankings by the median of the eleven ranks is obvious. With only a few measures a small difference in the median score resulting from chance error or the inherent

¹ Thorndike, E. L., *Mental and Social Measurements*, p. 225.

² *Ibid.*, p. 193.

lack of fine discriminations on account of the small number of tests, affects the median rank of the individual much more than several such differences affect the average rank. In the latter case such differences tend to offset each other or if they do not entirely balance each other they enter into a composite where so much does not depend upon a single measure.

One other point should be brought out in connection with the coefficients of correlation in Table XXXIV. By each of the three methods the correlation between the February 1916 and June 1917 rankings is about 10 per cent lower than the correlation between the rankings of the testings closer together in point of time. It has been found in another section of the study that the number of pupils making erratic scores and the number of erratic scores per pupil increased with each succeeding testing. Granting that there was improvement in all the abilities tested this shows that the amounts of improvement of different pupils in their different abilities were increasingly disproportionate the longer the pupils remained in school. The coefficients of correlation mentioned above tend to show that the improvement of the pupils in composite ability also was made at varying rates, and that the rate of improvement of different pupils did not fluctuate thus overcoming the inequalities, but rather that the inequalities became more pronounced the longer the pupils remained in school.

The practical importance of such varying rates of improvement bears upon the length of time an evaluation of an achievement by such tests can be considered as a valid index of the ability of the pupil. As examples of such variation two cases from this investigation are cited. Pupil No. 51 ranked 67 by the tests in February, 1916; 35 in February, 1917; and 8 in June, 1917. This pupil was placed in group 6 in school in February, 1916. By the judgment of the teachers he was advanced to group 5 in April, group 4 in May, and to group 3 in June, 1916. In February, 1917 he was in group 2 and in June, 1917 he was in group 1. Pupil No. 7 ranked 28 by the tests in February, 1916; 39 in February, 1917; and 63 in June, 1917. In February, 1916 this pupil was placed in group 3, in June, 1916, he was in group 4, in February, 1917, in group 5, and in June, 1917 he was still in group 5.

2. CORRELATION BETWEEN MEASURES OF VARIABILITY

Is there any constancy in the variability of the individual's achievement? This question can be studied by finding the amount of correlation between measures of variability in the different testings. Two methods of ranking the pupils are used. One method of ranking is by the extent of their entire range in the eleven tests. The individual having the smallest range in multiples of Q was ranked one, least variable, and the pupil having the largest range was ranked seventy-two, most variable. The other method is by the approximation of the interquartile range. The ranking was made in the same manner, the least variable being ranked one.

The pupils were ranked according to variability in each of the three testings by both of these methods. The three combinations of the rankings by each method were then correlated. The coefficients are given in Part A of Table XXXV. The results show consistently a small positive relation between the amount of variability in the three testings. Both methods show about the same results.

TABLE XXXV
CORRELATION BETWEEN MEASURES OF VARIABILITY IN THE ELEVEN
TESTS AT THE DIFFERENT TIMES THEY WERE GIVEN

		P.E. of r
(A)	Range in Values of Q	
	Feb. 1916 with Feb. 1917.....	.32..... .07
	Feb. 1917 with June 1917.....	.27..... .07
	Feb. 1916 with June 1917.....	.17..... .08
	Interquartile Range in Values of Q	
	Feb. 1916 with Feb. 1917.....	.20..... .08
	Feb. 1917 with June 1917.....	.29..... .07
	Feb. 1916 with June 1917.....	.18..... .08
(B)	Range in Values of Q	
	Feb. 1917 with Teachers' Ratings....	.20.... .08
	June 1917 with Teachers' Ratings....	.20.... .08
	Interquartile Range in Values of Q	
	June 1917 with Teachers' Ratings....	.22.... .07

In Part B of Table XXXV three of the rankings of Part A are used to correlate with the teachers' ratings in variability. These were secured as follows. Each teacher rated every pupil he or she had had in class as to the character of the work done. Under one of the three headings,—consistent, variable, and erratic—the teacher was asked to “check either the character of the

work in general or the character of the work in each subject." Variable was to be considered as the step between consistent and its opposite, erratic. From six to eleven ratings were thus secured for each pupil. These were turned into per cents of consistent, variable, and erratic ratings. The percentage of consistent was weighted by three, the percentage of variable by two, and the percentage of erratic by one. From the totals of these weighted percentages the pupils were ranked from one to seventy-two. The largest percentage was ranked one, least variable, and the smallest seventy-two, most variable.

The coefficients in Part B resulted from correlating these rankings with the rankings made from the range in the eleven tests. Here again, although not high, the coefficients show consistently a small amount of positive relation.

3. CORRELATION BETWEEN MEASURES OF ABILITY AND VARIABILITY

Having studied the resemblance between measures of ability and the resemblance between measures of variability the question naturally follows: What is the relation between ability and variability? The results from such correlations are shown in Table XXXVI.

TABLE XXXVI

CORRELATION BETWEEN MEASURES OF ABILITY AND VARIABILITY IN THE
ELEVEN TESTS. (HIGHEST ABILITY AND LEAST VARIABILITY
RANKED ONE.)

	<i>Feb.</i> 1916	<i>Feb.</i> 1917	<i>June</i> 1917	
Ability by Median Rank in Values of Q with Variability by Range in Values of Q	.19	.31	.33	
Ability by Median Rank in Values of Q with Variability by Inter-Quartile Range in Q	.26	.45	.43	
Composite Ability by Average Rank in Three Testings with Composite Variability by Range in Three Testings				.39
Composite Ability by Average Rank in Three Testings with Composite Variability by Inter-Quartile Range (ap- proximation) in Three Testings				.55

The rankings used in the two preceding topics were used to find the correlation between the ability and the variability of these pupils. Variability is correlated with ability for each of the three testings, first, by using the entire range in values of *Q* as the measure of variability, and second, by using the approximation of the interquartile range as the measure of variability. The median rank in values of *Q* is used as the measure of ability in both cases. The pupil having the highest median score is ranked one in ability and the pupil having the smallest range is ranked one in variability. The correlations result in positive coefficients in all cases, and interestingly, in greater amounts of relation when the coefficients secured in the later testings are compared with the coefficients of the first testing.

From the relations shown by the coefficients of correlation in this section the following summary may be made. Higher coefficients of correlation were obtained by ranking these pupils by their average achievement than by ranking them by their median achievement. When the number of tests given is relatively small the median is affected much more by slight deviations than is the average. The teachers' ratings in variability show positive coefficients when correlated with the variability as shown by the tests. The relation between ability and variability as expressed by coefficients of correlation is not great but is consistently positive. It was greater in the later testings than in the first testing.

VIII

CONCLUSIONS

The questions asked in connection with the statement of the problem may be grouped under four headings. Although all of these questions have not been fully answered, the following conclusions seem to be justified in view of the results obtained by testing at three different times, during a period of a year and a half, seventy-two junior high school boys with a group of eleven standardized scales and tests.

A. Concerning methods of comparing or equating individual measures of achievement.

The method of comparing the scores of an individual by ranks from highest to lowest in a group is not satisfactory for the purpose of diagnosing individual achievements. By this method much of the refinement of the original measures is lost. The method of transmuting the original scores into multiples of a measure of variability of the group produces more reliable results because practically all of the refinement of the original measures is preserved. The semi-interquartile-range or the average deviation is to be preferred to the mean square deviation as a measure of variability for this kind of statistical treatment as the latter weights too heavily the extreme and erratic scores.

B. Concerning the amount and distribution of individual variability.

The variability of the individual in these tests is a large fraction of the variability of the group. The average amount of individual variability, measured in terms of the Q, is 82 per cent of the group variability. This is evidence of the unreliability of one or a few tests for the purpose of educational prognosis.

The tests used in the second and third testings are not in all cases repetitions of the same tests or tests comparable in the amount of absolute variability, but the results of certain tests which are comparable tend to show that the absolute amount of group variability is about the same in all testings. The individual variability in terms of the group variability is the same in the first and second testings, and slightly greater in the third.

The form of distribution of the achievements of the individual approximates the normal surface of frequency. The mode is distinctly pronounced. The chief divergence from the normal curve is skewness downward from the median.

The average range in the achievements of the individual is 4.78Q in terms of the Q of the group. The average range in achievements above the individual medians is 1.94Q, and the average range below the individual medians is 2.84Q.

The lowest ranking pupils are the most variable in their achievements. The variability of the second tertile is greater than the variability of the first tertile, and the variability of the third tertile is greater than that of the second tertile in each testing. Measured by the Q of the group the average of the three testings shows that the variability of the highest tertile is .69Q, that of the middle tertile, .78Q, and that of the lowest tertile, .99Q.

The overlappings of the divisions of the group show marked amounts of difference between the median achievement of the different quartiles. In terms of the Q of the group the median achievement of the second quartile is .95Q lower than that of the first; that of the third, .65Q lower than that of the second; and the median achievement of the fourth is .84Q lower than that of the third. Therefore the pupils of this group who are most variable in their achievements are also distinctly lowest in achievements as measured by these tests.

For the purpose of individual diagnosis it would be of advantage to have more tests scaled from the zero point and standardized in variability either by grade or by age of the pupil.

C. Concerning extremely variable or erratic scores.

Considering as erratic all scores at a distance of 3Q or more in each direction from the median score of the individual the average number of erratic scores for each testing is 4.5 per cent of the total number of scores for each testing. Twenty-nine per cent of the erratic scores are plus and 71 per cent are minus.

Spelling caused more erratic scores than any other test with the exception of Algebra, Addition and Subtraction, which was in process of construction and which was given only once. In spelling and the three rate tests,—opposites, mixed relations, and easy directions—all but one of the erratic scores are minus.

These four tests contain 46 per cent of the erratic scores. In the remaining seven tests the total number of erratic scores plus is 30 and the total number minus is 29.

There are 108 erratic scores in the three testings,—24 per cent in the first, 35 per cent in the second, and 41 per cent in the third testing. In the first testing 35 per cent of the boys made one or more erratic scores; in the second testing, 42 per cent; and in the third testing, 47 per cent. The distribution of erratic scores among the tertiles is as follows: 22 per cent are in the first tertile; 29 per cent are in the second; and 49 per cent are in the third tertile. Using the average number of boys who made erratic scores it is found that 24 per cent are in the first tertile; 32 per cent are in the second; and 44 per cent are in the third tertile. Therefore the results of this study show a noticeable increase in the number of pupils making erratic scores in the later testings and a slight increase in the number of erratic scores per pupil in the second and third testings.

In this group of pupils 76 per cent made one or more erratic scores. Forty-two per cent made erratic scores in one testing only; 22 per cent in two testings; and 12 per cent in all three testings. Of the pupils who made erratic scores 55 per cent made them in the minus direction only; 16 per cent in the plus direction only; and 29 per cent in both the plus and minus directions. No distinct types of variation are found in this group of pupils.

A re-examination under closely controlled conditions of a few boys who made the most variable scores in spelling and in the rate tests produced an average reduction of 25 per cent in the number of erratic scores on the basis of the total number of scores in the re-examination.

Five possible factors in the causation of erratic scores were studied. They are: the nature of the tests used, the administration of the tests, accidental or unusual occurrences, statistical treatment of the results, and the ability of the individual in different traits. The nature of the tests and the statistical treatment of the results seem to have magnified the variability of a relatively small number of scores. The administration of the tests in so far as it can be judged was an unimportant factor. Accidental or unusual occurrences probably caused a small pro-

portion of the erratic scores. From the evidence of this study it appears that the ability of the pupil in different traits was the greatest factor in the causation of scores that varied 3Q or more from the individual's median.

D. Concerning the relation between measures of ability, between measures of variability, and between measures of ability and variability.

The coefficients of correlation between the different testings show that from the results of the first testing the average achievement of these boys in similar tests a year later and a year and a half later could be predicted with a rather high degree of accuracy. However, the results of the tests and the judgments of the teachers agree in showing a very great amount of change in the ranking of certain individuals among the group in the later testings. For the purpose of individual diagnosis the results obtained from a single testing with such a group of tests should be considered as indices of individual ability which will be valid for varying lengths of time. Such results should be supplemented and checked by repetitions of the same or similar tests. School organization should be flexible enough to allow for a shifting among groups for instruction commensurate with the relative gain or loss in ability on the part of certain individuals.

The correlation between the first and third testings which are a year and a half apart in point of time is about 10 per cent less than the correlation between either the first and second or the second and third testings. This supplements the evidence already found showing that the pupils vary more in their achievements the longer they remain in school.

The amount of correlation between measures of variability in the different testings, although small, is positive in all cases.

The coefficient of correlation between composite ability by average rank in the three testings and composite variability by interquartile range (approximation) in the three testings is .55. This seems to indicate that there was a considerable amount of relation between the ability of these pupils to achieve in these tests and the consistency or lack of variability in their achievements.

APPENDIX

TABLE XXXVII

DISTRIBUTION OF SCORES ABOVE AND BELOW THE INDIVIDUAL MEDIANS IN THE ELEVEN TESTS TRANSMUTED INTO MULTIPLES OF Q BY THE ORIGINAL DISTRIBUTIONS

The Measure of Central Tendency is the Median of the Individual's Scores Transmuted Into Multiples of Q.

Value in Q	Feb. 1916				Feb. 1917				June 1917			
	Tertile			Total	Tertile			Total	Tertile			Total
	I	II	III		I	II	III		I	II	III	
+5.0 to —											1	1
+4.5 to +4.9											1	1
+4.0 to +4.4		1		1						1	1	2
+3.5 to +3.9					1	1		2	1		2	3
+3.0 to +3.4	2	3	2	7	3	6		9	1	1	3	5
+2.5 to +2.9	2	4	5	11	2	2	2	6	3	2	2	7
+2.0 to +2.4	4	10	6	20	5	7	9	21	3	7	14	24
+1.5 to +1.9	13	14	19	46	8	9	18	35	6	9	21	36
+1.0 to +1.4	16	24	33	73	21	20	26	67	16	31	25	72
+ .5 to + .9	37	36	30	103	38	34	35	107	49	33	31	113
0 to + .4	58	40	37	135	58	56	35	149	53	48	31	132
0 to — .4	45	42	49	136	50	47	45	142	51	47	41	139
— .5 to — .9	33	32	31	96	33	29	19	81	29	33	35	97
—1.0 to —1.4	18	23	17	58	16	25	23	64	21	17	17	55
—1.5 to —1.9	16	16	15	47	17	15	12	44	12	11	12	35
—2.0 to —2.4	10	7	5	22	7	6	12	25	5	7	6	18
—2.5 to —2.9	5	6	8	19	2	5	6	13	6	7	7	20
—3.0 to —3.4	3		3	6	1	2	2	5	3	4	4	11
—3.5 to —3.9	1	2	1	4			4	4	3	2	1	6
—4.0 to —4.4	1			1	2	1		3	2	2	2	6
—4.5 to —4.9		1	1	2	2	1	3	6			3	3
—5.0 to —		3	2	5	2	1	6	9		2	4	6
No. of Scores	264	264	264	792	264	264	264	792	264	264	264	792

TABLE XXXVIII

DISTRIBUTION OF SCORES ABOVE AND BELOW THE INDIVIDUAL MEDIANS IN
CERTAIN TESTS TRANSMUTED INTO MULTIPLES OF Q BY
THE ORIGINAL DISTRIBUTIONS

*The Measure of Central Tendency is the Median of the Individual's Scores
Transmuted Into Multiples of Q.*

<i>Value in Q</i>	<i>Eight Trabue</i>	<i>Six Mathe- matics</i>	<i>Five Directions</i>	<i>Three Reading</i>
+5.0 to —				
+4.5 to +4.9				
+4.0 to +4.4				
+3.5 to +3.9	2		1	
+3.0 to +3.4	6	2		
+2.5 to +2.9	12	5	2	2
+2.0 to +2.4	15	13	5	3
+1.5 to +1.9	20	16	17	7
+1.0 to +1.4	55	34	28	9
+ .5 to + .9	62	61	40	23
.0 to + .4	112	85	87	64
.0 to — .4	109	80	76	55
— .5 to — .9	73	63	37	19
—1.0 to —1.4	51	28	34	13
—1.5 to —1.9	29	15	17	7
—2.0 to —2.4	7	13	8	10
—2.5 to —2.9	9	8	2	1
—3.0 to —3.4	4	4	3	1
—3.5 to —3.9			1	1
—4.0 to —4.4		1		1
—4.5 to —4.9	2	2	1	
—5.0 to —		2	1	
No. of Scores	568 *	432	350	216

* Scores are lacking for four individuals in both Trabue D and E.

TABLE XXXIX

ORIGINAL SCORES BY TESTS AND BY INDIVIDUALS. FEBRUARY, 1916

<i>Individual</i>	<i>Woody Mult.</i>	<i>Woody Div.</i>	<i>Trabue B</i>	<i>Trabue C</i>	<i>Reading Alpha 2</i>	<i>Visual Vocabulary</i>	<i>Composition</i>	<i>Spelling</i>	<i>Opposites</i>	<i>Mixed Relations</i>	<i>Easy Directions</i>
1.....	32	33	12	11	20	26	30.2	48	13	2	9
2.....	32	33	17	11	26	42	23.4	48	18	18	17
3.....	31	22	14	13	23	26	22.2	49	17	10	11
4.....	35	33	16	16	25	33	48.6	48	18	18	16
5.....	33	26	11	12	23	24	27.5	47	13	5	19
6.....	35	33	11	15	27	26	24.5	50	18	19	11
7.....	30	30	14	13	21	28	37.8	49	14	19	17
8.....	31	29	12	6	21	30	37.7	48	18	16	19
9.....	32	28	13	14	29	32	44.7	46	17	19	20
10.....	35	30	11	11	25	31	40.6	48	14	18	16
11.....	37	32	13	15	26	34	43.2	49	10	16	13
12.....	33	31	11	13	21	20	36.	47	12	9	20
13.....	29	24	4	10	6	22	29.2	44	14	8	12
14.....	32	32	11	14	25	28	27.4	49	19	13	20
15.....	31	27	7	10	14	15	29.8	40	12	2	9
16.....	28	29	10	10	31	30	34.5	48	17	6	15
17.....	35	31	11	14	22	29	33.6	50	17	17	20
18.....	28	25	14	14	26	36	34.8	45	16	14	16
19.....	21	24	14	10	22	15	33.2	41	16	3	10
20.....	38	32	14	12	22	25	27.4	45	13	15	11
21.....	34	28	17	18	32	40	39.2	50	19	20	16
22.....	31	26	9	11	13	16	25.2	29	14	9	11
23.....	28	27	12	13	30	34	31.	42	16	18	17
24.....	35	31	12	13	25	42	34.	50	17	9	13
25.....	27	32	12	14	24	35	26.5	45	19	19	19
26.....	31	32	12	13	30	36	50.	49	19	19	14
27.....	31	30	16	12	26	29	37.5	37	19	13	13
28.....	35	34	13	11	19	24	39.6	47	16	4	13
29.....	32	16	16	15	30	28	32.	49	14	16	13
30.....	29	27	12	15	32	31	43.5	49	18	18	12
31.....	35	34	12	10	19	22	35.2	49	15	18	11
32.....	34	34	17	12	19	19	33.	48	17	13	13
33.....	32	33	14	12	34	30	32.2	49	19	5	16
34.....	30	31	15	15	28	38	21.2	49	18	20	18
35.....	30	33	12	12	22	28	30.	40	17	16	13
36.....	30	27	14	13	19	31	33.5	50	15	11	11
37.....	22	25	11	11	16	29	28.	48	19	18	16
38.....	36	33	15	16	30	33	44.5	48	19	5	18
39.....	34	31	10	11	26	20	25.2	46	13	13	13
40.....	31	25	16	10	33	27	28.5	44	17	10	14

TABLE XXXIX—*Continued*

ORIGINAL SCORES BY TESTS AND BY INDIVIDUALS. FEBRUARY, 1916

<i>Individual</i>	<i>Woody Mult.</i>	<i>Woody Div.</i>	<i>Trabue B</i>	<i>Trabue C</i>	<i>Reading Alpha 2</i>	<i>Visual Vocabulary</i>	<i>Composition</i>	<i>Spelling</i>	<i>Opposites</i>	<i>Mixed Relations</i>	<i>Easy Directions</i>
41.....	25	24	10	12	22	31	30.8	42	18	7	16
42.....	31	31	15	10	31	32	30.6	49	16	11	18
43.....	34	31	15	11	22	28	40.6	49	18	17	17
44.....	33	23	7	12	21	26	43.5	43	18	16	10
45.....	27	28	14	18	34	34	44.	48	19	18	19
46.....	37	32	8	10	14	28	26.4	48	16	10	13
47.....	36	32	13	15	29	36	40.8	49	20	17	14
48.....	33	29	13	16	27	39	46.	45	41	19	14
49.....	37	34	10	11	15	31	38.5	48	18	9	7
50.....	32	29	12	13	21	34	42.5	50	17	16	14
51.....	11	31	10	12	26	30	17.5	44	14	13	12
52.....	31	27	15	12	35	37	56.6	49	16	5	14
53.....	32	27	13	12	29	30	47.8	46	15	5	11
54.....	35	27	10	10	20	26	34.	48	15	7	11
55.....	31	32	11	10	20	35	40.8	47	19	8	18
56.....	27	27	12	13	23	24	34.8	27	18	17	12
57.....	29	23	13	11	32	29	36.6	46	20	19	13
58.....	30	26	8	13	23	32	34.2	49	12	17	19
59.....	33	26	12	13	29	38	38.8	46	17	19	14
60.....	33	33	16	11	28	27	33.4	50	15	6	14
61.....	33	34	10	12	24	25	28.5	48	10	13	13
62.....	35	28	15	15	25	33	22.6	47	18	15	11
63.....	36	28	13	12	21	23	32.6	49	19	4	14
64.....	36	30	15	16	30	41	54.	49	20	20	20
65.....	31	26	13	10	20	20	23.7	30	17	11	11
66.....	35	29	8	13	16	16	36.6	47	15	11	8
67.....	33	31	12	12	28	32	35.8	48	18	15	15
68.....	33	31	9	15	14	20	26.5	48	11	11	10
69.....	27	27	12	11	17	30	26.7	46	15	17	14
70.....	33	31	11	12	19	31	20.2	49	18	16	19
71.....	29	33	10	12	18	28	34.5	46	11	4	16
72.....	24	31	11	9	19	26	30.4	49	14	4	9
73.....	33	33	13	11	20	19	26.	46	19	14	15
74.....	33	33	17	11	38	24	31.7	48	17	11	14
75.....	33	31	14	13	22	32	46.5	49	18	4	16

TABLE XL

ORIGINAL SCORES BY TESTS AND BY INDIVIDUALS. FEBRUARY, 1917

<i>Individual</i>	<i>Woody Mult.</i>	<i>Woody Div.</i>	<i>Trabue J</i>	<i>Trabue K</i>	<i>Reading Alpha 2</i>	<i>Visual Vocabulary</i>	<i>Composition</i>	<i>Spelling</i>	<i>Opposites</i>	<i>Mixed Relations</i>	<i>Easy Directions</i>
1.....	16	13	5	8	22	97	52.	43	16	10	14
2.....	15	13	6	8	29	126	44.7	46	20	2	17
3.....	14	13	6	5	25	119	51.2	34	16	12	19
4.....	19	14	9	12	37	122	60.5	45	20	17	18
5.....	17	12	5	2	38	131	39.2	42	19	8	15
6.....	13	14	8	5	32	114	43.5	39	20	17	19
7.....	15	12	8	6	33	108	58.2	39	20	9	20
8.....	17	14	7	10	40	133	43.7	48	20	18	19
9.....	17	11	7	8	38	142	53.5	49	20	17	16
10.....	16	14	5	3	35	125	45.2	37	20	15	16
11.....	14	14	6	7	28	133	55.5	40	20	13	16
12.....	14	14	8	6	29	100	41.5	41	20	19	20
13.....	11	12	5	5	16	51	41.	38	19	3	17
14.....	16	13	4	6	31	109	44.5	45	20	10	20
15.....	17	12	4	5	13	87	42.	39	18	6	14
16.....	16	13	6	6	26	108	47.5	40	20	13	18
17.....	16	13	8	9	37	117	53.7	47	20	17	20
18.....	14	9	8	8	39	126	46.	38	20	17	20
19.....	12	11	7	5	15	100	41.5	27	17	6	11
20.....	15	13	8	5	25	123	38.5	44	18	13	20
21.....	18	13	10	9	37	162	42.	47	20	18	20
22.....	16	13	4	6	30	35	46.	20	16	15	12
23.....	14	13	8	10	33	140	47.5	33	20	12	18
24.....	16	14	8	8	29	130	49.7	48	20	5	20
25.....	14	13	7	5	30	116	50.5	40	19	10	20
26.....	18	14	7	8	39	144	45.2	47	20	20	20
27.....	13	13	8	12	36	163	45.	35	20	17	20
28.....	16	14	8	10	27	106	49.	36	20	18	19
29.....	16	7	6	8	27	146	47.	39	20	19	19
30.....	15	13	10	10	37	120	48.	44	20	5	19
31.....	17	13	9	4	26	100	56.5	44	20	18	18
32.....	19	13	5	6	23	106	48.2	50	20	15	17
33.....	17	14	7	4	39	121	58.7	42	20	18	20
34.....	11	12	8	7	31	129	47.5	44	20	19	18
35.....	15	13	8	8	30	97	40.5	23	20	17	16
36.....	16	11	8	6	20	106	38.2	45	20	7	13
37.....	14	13	4	5	31	96	39.5	48	20	17	19
38.....	17	14	7	8	42	139	58.7	44	20	12	17
39.....	17	12	4	6	35	117	47.2	42	19	12	11
40.....	12	13	9	6	36	117	54.5	37	20	16	18

TABLE XL—Continued

ORIGINAL SCORES BY TESTS AND BY INDIVIDUALS. FEBRUARY, 1917

<i>Individual</i>	<i>Woody Mult.</i>	<i>Woody Div.</i>	<i>Trabue J</i>	<i>Trabue K</i>	<i>Reading Alpha 2</i>	<i>Visual Vocabulary</i>	<i>Composition</i>	<i>Spelling</i>	<i>Opposites</i>	<i>Mixed Relations</i>	<i>Easy Directions</i>
41.....	14	11	4	7	30	118	43.2	27	20	16	20
42.....	15	13	7	8	37	145	43.2	46	20	16	20
43.....	15	14	9	7	34	125	45.7	44	20	19	20
44.....	16	12	6	7	38	136	44.	37	20	11	19
45.....	14	14	10	10	36	143	60.5	40	20	19	19
46.....	14	13	6	7	31	123	47.2	40	20	18	20
47.....	16	12	6	8	35	142	47.2	44	20	17	20
48.....	17	13	7	12	39	144	53.	41	20	19	19
49.....	17	12	8	4	20	117	44.2	45	20	15	20
50.....	15	14	6	6	29	136	48.7	48	20	16	17
51.....	15	13	6	6	36	111	57.7	46	20	10	19
52.....	14	10	6	8	38	158	56.	46	20	16	18
53.....	16	13	5	5	32	121	42.	45	20	8	18
54.....	16	13	6	7	29	118	39.	35	16	5	14
55.....	15	13	10	9	28	129	45.7	47	19	12	20
56.....	15	12	6	5	29	115	42.	16	17	12	17
57.....	14	11	11	6	32	123	48.7	44	19	17	20
58.....	18	13	8	7	20	113	42.	48	20	17	19
59.....	15	13	8	6	33	144	50.5	38	20	17	20
60.....	16	13	9	8	34	142	37.7	46	20	18	20
61.....	16	13	6	7	35	129	39.7	46	11	19	17
62.....	17	13	6	5	17	104	42.7	43	19	19	12
63.....	15	13	7	5	34	133	49.2	45	20	16	18
64.....	15	14	11	10	41	154	52.7	46	20	20	20
65.....	13	14	5	4	25	103	48.	36	19	16	18
66.....	14	13	6	6	22	109	47.	47	17	7	7
67.....	14	14	9	5	34	120	55.5	44	19	13	19
68.....	14	14	5	3	25	112	41.7	33	18	16	20
69.....	15	12	7	6	19	137	40.	39	20	18	20
70.....	18	12	8	5	29	140	48.2	49	20	17	16
71.....	12	14	5	6	23	73	44.2	35	20	17	20
72.....	9	12	7	5	29	106	40.	40	18	12	15
73.....	17	13	6	4	31	87	43.2	33	20	18	17
74.....	17	12	7	5	37	138	45.	47	20	19	19
75.....	14	12	10	9	32	118	51.2	49	20	14	20

TABLE XLI
ORIGINAL SCORES BY TESTS AND BY INDIVIDUALS. JUNE, 1917

<i>Individual</i>	<i>Algebra Add. Subt.</i>	<i>Algebra Mult. Div.</i>	<i>True L</i>	<i>True M</i>	<i>Reading</i>	<i>Visual Vocabulary</i>	<i>Composition</i>	<i>Spelling</i>	<i>Opposites</i>	<i>Mixed Relations</i>	<i>Easy Directions</i>
1.....	12	13	3	6	51	97	51.3	49	15	14	20
2.....	4	7	6	10	55	109	52.3	44	15	18	19
3.....	10	8	6	5	52	105	50.	43	14	12	20
4.....											
5.....	10	8	9	9	51	101	47.	46	15	10	19
6.....	10	10	3	14	51	108	51.3	45	19	20	20
7.....	8	8	4	9	52	99	51.6	41	16	13	19
8.....	19	17	6	7	48	99	60.6	50	19	18	20
9.....	9	9	4	8	55	107	60.3	50	20	19	18
10.....	13	8	5	9	49	108	53.3	43	12	18	18
11.....	18	15	6	7	51	110	54.	48	18	20	19
12.....	11	14	5	4	45	96	51.	48	18	20	20
13.....	9	10	4	3	30	63	52.	35	17	1	17
14.....	18	13	5	4	53	101	46.	45	14	15	20
15.....	15	12	4	1	27	78	43.	46	16	4	16
16.....	5	8	6	4	51	101	52.	42	19	13	18
17.....	12	15	7	9	56	109	67.6	48	18	16	20
18.....	7	7	11	8	47	106	62.	44	19	10	18
19.....	7	8	4	7	35	77	51.6	38	11	9	14
20.....	10	12	5	4	51	99	46.	41	17	19	20
21.....	11	16	6	12	56	111	62.	49	19	20	20
22.....											
23.....	9	9	8	9	50	109	57.3	35	20	20	20
24.....	10	12	7	7	43	109	64.3	47	19	0	19
25.....	9	7	5	8	55	104	63.6	41	17	16	20
26.....	16	15	10	9	52	115	61.6	46	19	20	19
27.....	10	6	7	12	46	110	60.6	48	18	16	20
28.....	11	12	7	7	47	102	47.3	43	20	19	20
29.....	7	17	7	7	53	113	50.6	42	18	8	19
30.....	9	12	7	8	48	112	69.6	49	15	19	17
31.....	12	15	2	7	48	97	67.	45	18	18	19
32.....	11	11	3	4	39	65	51.	50	16	12	19
33.....	11	12	5	9	51	108	56.6	44	19	20	20
34.....	14	11	6	11	50	104	58.3	48	18	18	19
35.....	9	11	5	4	48	90	49.3	33	18	17	18
36.....	11	7	7	5	47	112	52.3	48	16	13	15
37.....											
38.....	12	16	8	10	55	108	65.3	48	19	18	20
39.....	12	14	6	2	46	108	46.	41	14	11	19
40.....	11	11	6	4	44	117	53.	46	17	15	20

TABLE XLI—*Continued*
 ORIGINAL SCORES BY TESTS AND BY INDIVIDUALS. JUNE, 1917

<i>Individual</i>	<i>Algebra Add. Subt.</i>	<i>Algebra Mult. Div.</i>	<i>Trabue L</i>	<i>Trabue M</i>	<i>Reading</i>	<i>Visual Vocabulary</i>	<i>Composition</i>	<i>Spelling</i>	<i>Opposites</i>	<i>Mixed Relations</i>	<i>Easy Directions</i>
41.....	9	4	4	5	43	107	48.	35	16	18	20
42.....	10	10	7	8	53	110	49.3	47	16	17	19
43.....	10	5	6	9	45	105	51.3	45	18	20	20
44.....	10	14	4	4	47	102	52.3	33	18	18	18
45.....	14	11	8	9	55	112	65.	49	19	18	20
46.....	17	17	4	2	52	101	48.3	40	15	16	19
47.....	17	16	3	11	55	110	60.	49	16	17	18
48.....	14	16	7	8	52	110	51.6	50	13	20	19
49.....	11	14	6	4	44	105	53.6	45	19	13	19
50.....	8	10	7	7	50	104	48.3	50	18	15	18
51.....	15	19	8	10	50	112	59.6	49	18	15	19
52.....	5	7	11	9	53	115	67.3	48	19	15	20
53.....	12	9	6	6	55	105	53.	49	17	16	20
54.....	14	12	5	5	37	106	44.	39	15	6	16
55.....	9	11	6	5	55	98	57.6	47	19	19	20
56.....	15	15	5	5	49	107	57.6	25	17	17	17
57.....	11	12	7	10	44	108	60.3	49	17	18	19
58.....	14	11	6	8	45	111	57.	48	19	14	20
59.....	11	10	5	6	53	108	59.6	47	16	19	19
60.....	16	17	9	13	56	113	59.	47	17	11	20
61.....	11	16	5	3	51	115	44.3	44	13	20	18
62.....	15	17	5	5	47	107	42.6	43	18	19	20
63.....	12	16	5	9	52	112	67.	50	16	16	17
64.....	11	15	7	11	53	110	61.3	48	19	18	20
65.....	13	11	7	5	53	97	53.	33	14	18	19
66.....	11	15	5	6	37	107	45.3	46	12	12	11
67.....	11	13	6	6	51	92	54.3	47	18	15	20
68.....	17	15	7	7	53	106	44.6	36	17	19	20
69.....	7	6	6	9	43	108	53.	48	18	8	20
70.....	10	10	7	7	51	104	65.6	50	16	17	19
71.....	18	11	4	5	38	99	57.3	42	15	11	19
72.....	13	12	5	7	52	110	46.	46	17	5	17
73.....	8	11	6	5	52	103	58.3	32	19	19	19
74.....	10	15	6	8	56	108	57.3	44	17	16	19
75.....	9	11	11	9	52	105	54.	50	15	7	20

TABLE XLII

ORIGINAL SCORES BY TESTS AND BY INDIVIDUALS. ADDITIONAL TESTS

<i>Individual</i>	<i>Trabue D</i>	<i>Trabue E</i>	<i>Hard Directions</i>	<i>Hard Directions</i>	<i>Individual</i>	<i>Trabue D</i>	<i>Trabue E</i>	<i>Hard Directions</i>	<i>Hard Directions</i>
1.....	11	11	21	18	41.....	14	14	21	20
2.....	14	17	20	22	42.....	14	13	22	22
3.....			22	21	43.....	17	15	18	19
4.....			20		44.....	14	14	21	18
5.....	12	13	17	18	45.....	15	19	20	22
6.....	18	13	18	17	46.....	14	14	16	17
7.....	16	12	22	21	47.....	14	12	20	20
8.....	13	12	16	17	48.....	15	12	21	19
9.....	15	13	19	21	49.....	12	16	19	19
10.....	14	17	21	18	50.....	13	13	14	18
11.....	14	16	21	22	51.....	13	11	19	19
12.....	17	12	18	20	52.....	15	18	20	21
13.....	11	13	16	14	53.....	12	8	22	19
14.....	16	12	15	19	54.....	11	13	13	13
15.....	9	10	11	13	55.....	16	17	18	19
16.....	15	11	16	17	56.....	16	16	18	20
17.....	15	17	17	16	57.....			21	20
18.....	17	14	22	21	58.....	10	10	21	19
19.....	14	12	10	14	59.....			22	21
20.....	13	14	17	19	60.....	15	17	22	22
21.....	16	15	22	22	61.....	11	14	14	15
22.....	11	8	15		62.....	14	14	12	15
23.....	14	15	20	20	63.....	14	14	12	17
24.....	14	16	22	21	64.....	17	16	22	21
25.....	16	13	15	19	65.....	12	12	15	17
26.....	15	17	18	21	66.....	10	13	12	13
27.....	14	15	21	22	67.....	13	14	20	21
28.....	15	13	15	15	68.....	13	14	22	20
29.....	15	17	19	22	69.....			16	20
30.....	14	18	16	20	70.....	15	15	20	17
31.....	13	16	19	20	71.....	12	14	18	18
32.....	14	13	18	19	72.....	15	13	20	20
33.....	13	14	22	20	73.....	19	15	16	20
34.....	12	16	21	21	74.....	16	13	20	22
35.....	13	12	18	21	75.....	11	8	16	18
36.....	16	16	13	19					
37.....	12	14	17						
38.....	17	17	21	21					
39.....	14	13	15	22					
40.....	16	13	19	21					

VITA

The writer of this dissertation, CHESTER ARTHUR BUCKNER, was born near Highland Center, Iowa, January 28, 1885. He attended the village school at Highland Center intermittently until 1900 when he received a common school diploma from Wapello County, Iowa. He entered the high school at Ottumwa, Iowa, in 1901 and was graduated in 1905. His undergraduate work, and part of his graduate work, was done at the State University of Iowa, Iowa City, Iowa, where he received the degree of Bachelor of Arts in 1909 and the degree of Master of Arts in 1911. He was a student at Teachers College for a year and a half beginning February, 1916. The writer's teaching experience includes one year as teacher of mathematics and civics in the high school at Clinton, Iowa; two years as head of the department of English in the high school at Manila, Philippine Islands; one year and one-half and two summer sessions as assistant professor of education in the University of Kansas; and one year in charge of educational measurements in The Lincoln School of Teachers College.



